

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 May 2002 (23.05.2002)

PCT

(10) International Publication Number
WO 02/40720 A2

(51) International Patent Classification⁷: **C12Q 1/68**,
G01N 33/574, C07K 14/47, C12N 15/12, C07K 16/18

Yongming [US/US]; 869 S. Winchester Boulevard, Apartment 260, San Jose, CA 95128 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(21) International Application Number: PCT/US01/45010

(22) International Filing Date:
20 November 2001 (20.11.2001)

(74) Agents: **LICATA, Jane, Massey et al.**; Licata & Tyrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/249,997 20 November 2000 (20.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/249,997 (CIP)
Filed on 20 November 2000 (20.11.2000)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, ZA, ZW.

(71) Applicant (*for all designated States except US*): **DI-ADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SALCEDA, Susana** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US); **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US); **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US); **CAFFERKEY, Robert** [IE/US]; 350 Elan Village Lane, Apartment 218, San Jose, CA 95134 (US); **SUN,**

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/40720 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovary tissue, identifying ovary tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application
5 Serial No. 60/249,997 filed November 20, 2000, which is herein incorporated by
reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and
10 polypeptides present in normal and neoplastic ovary cells, including fragments, variants
and derivatives of the nucleic acids and polypeptides. The present invention also relates
to antibodies to the polypeptides of the invention, as well as agonists and antagonists of
the polypeptides of the invention. The invention also relates to compositions comprising
the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists
15 of the invention and methods for the use of these compositions. These uses include
identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and
non-cancerous disease states in ovary tissue, identifying ovary tissue and monitoring and
identifying and/or designing agonists and antagonists of polypeptides of the invention.
The uses also include gene therapy, production of transgenic animals and cells, and
20 production of engineered ovary tissue for treatment and research.

BACKGROUND OF THE INVENTION

Cancer of the ovaries is the fourth-most cause of cancer death in women in
the United States, with more than 23,000 new cases and roughly 14,000 deaths
predicted for the year 2001. Shridhar, V. et al., Cancer Res. 61(15):
25 5895-904 (2001); Memarzadeh, S. & Berek, J. S., J. Reprod. Med. 46(7):
621-29 (2001). The incidence of ovarian cancer is of serious concern
worldwide, with an estimated 191,000 new cases predicted annually.
Runnebaum, I. B. & Stickeler, E., J. Cancer Res. Clin. Oncol. 127(2): 73-79
(2001). Because women with ovarian cancer are typically asymptomatic until
30 the disease has metastasized, and because effective screening for ovarian
cancer is not available, roughly 70% of women present with an advanced stage

of the cancer, with a five-year survival rate of ~25-30% at that stage.

Memarzadeh, S. & Berek, J. S., *supra*; Nunns, D. et al., *Obstet. Gynecol.*

Surv. 55(12): 746-51. Conversely, women diagnosed with early stage ovarian

cancer enjoy considerably higher survival rates. Werness, B. A. &

- 5 Eltabbakh, G. H., *Int'l. J. Gynecol. Pathol.* 20(1): 48-63 (2001).

Although our understanding of the etiology of ovarian cancer is incomplete,

the results of extensive research in this area point to a combination of

age, genetics, reproductive, and dietary/environmental factors. Age is a

key risk factor in the development of ovarian cancer: while the risk for

- 10 developing ovarian cancer before the age of 30 is slim, the incidence of

ovarian cancer rises linearly between ages 30 to 50, increasing at a slower

rate thereafter, with the highest incidence being among septagenarian women.

Jeanne M. Schilder et al., *Hereditary Ovarian Cancer: Clinical Syndromes*

and Management, in *Ovarian Cancer* 182 (Stephen C. Rubin & Gregory P. Sutton

- 15 eds., 2d ed. 2001).

With respect to genetic factors, a family history of ovarian cancer is the

most significant risk factor in the development of the disease, with that

risk depending on the number of affected family members, the degree of their

relationship to the woman, and which particular first degree relatives are

- 20 affected by the disease. *Id.* Mutations in several genes have been

associated with ovarian cancer, including BRCA1 and BRCA2, both of which

play a key role in the development of breast cancer, as well as hMSH2 and

hMLH1, both of which are associated with hereditary non-polyposis ovary

cancer. Katherine Y. Look, *Epidemiology, Etiology, and Screening of Ovarian*

- 25 *Cancer*, in *Ovarian Cancer* 169, 171-73 (Stephen C. Rubin & Gregory P. Sutton

eds., 2d ed. 2001). BRCA1, located on chromosome 17, and BRCA2, located on

chromosome 13, are tumor suppressor genes implicated in DNA repair; mutations

in these genes are linked to roughly 10% of ovarian cancers. *Id.* at 171-72;

Schilder et al., *supra* at 185-86. hMSH2 and hMLH1 are associated with DNA

- 30 mismatch repair, and are located on chromosomes 2 and 3, respectively; it has

been reported that roughly 3% of hereditary ovarian carcinomas are due to

mutations in these genes. Look, *supra* at 173; Schilder et al., *supra* at

184, 188-89.

Reproductive factors have also been associated with an increased or reduced risk of ovarian cancer. Late menopause, nulliparity, and early age at menarche have all been linked with an elevated risk of ovarian cancer.

- 5 Schilder et al., *supra* at 182. One theory hypothesizes that these factors increase the number of ovulatory cycles over the course of a woman's life, leading to "incessant ovulation," which is thought to be the primary cause of mutations to the ovarian epithelium. *Id.*; Laura J. Havrilesky & Andrew Berchuck, *Molecular Alterations in Sporadic Ovarian Cancer*, in *Ovarian*
10 *Cancer* 25 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). The mutations may be explained by the fact that ovulation results in the destruction and repair of that epithelium, necessitating increased cell division, thereby increasing the possibility that an undesired mutation will occur. *Id.* Support for this theory may be found in the fact pregnancy,
15 lactation, and the use of oral contraceptives, all of which suppress ovulation, confer a protective effect with respect to developing ovarian cancer. *Id.*

- Among dietary/environmental factors, there would appear to be an association between high intake of animal fat or red meat and ovarian cancer, while the
20 antioxidant Vitamin A, which prevents free radical formation and also assists in maintaining normal cellular differentiation, may offer a protective effect. *Look, supra* at 169. Reports have also associated asbestos and hydrous magnesium trisilicate (talc), the latter of which may be present in diaphragms and sanitary napkins. *Id.* at 169-70.

- 25 Current screening procedures for ovarian cancer, while of some utility, are quite limited in their diagnostic ability, a problem that is particularly acute at early stages of cancer progression when the disease is typically asymptomatic yet is most readily treated. Walter J. Burdette, *Cancer: Etiology, Diagnosis, and Treatment* 166 (1998); Memarzadeh & Berek, *supra*;
30 Runnebaum & Stickeler, *supra*; Werness & Eltabbakh, *supra*. Commonly used screening tests include bimanual rectovaginal pelvic examination, radioimmunoassay to detect the CA-125 serum tumor marker, and transvaginal

ultrasonography. Burdette, *supra* at 166.

Pelvic examination has failed to yield adequate numbers of early diagnoses, and the other methods are not sufficiently accurate. *Id.* One study reported that only 15% of patients who suffered from ovarian cancer were diagnosed
5 with the disease at the time of their pelvic examination. Look, *supra* at 174. Moreover, the CA-125 test is prone to giving false positives in pre-menopausal women and has been reported to be of low predictive value in post-menopausal women. *Id.* at 174-75. Although transvaginal ultrasonography is now the preferred procedure for screening for ovarian
10 cancer, it is unable to distinguish reliably between benign and malignant tumors, and also cannot locate primary peritoneal malignancies or ovarian cancer if the ovary size is normal. Schilder et al., *supra* at 194-95. While genetic testing for mutations of the BRCA1, BRCA2, hMSH2, and hMLH1 genes is now available, these tests may be too costly for some patients and
15 may also yield false negative or indeterminate results. Schilder et al., *supra* at 191-94.

The staging of ovarian cancer, which is accomplished through surgical exploration, is crucial in determining the course of treatment and management of the disease. AJCC Cancer Staging Handbook 187 (Irvin D. Fleming et al. eds., 5th ed. 1998); Burdette, *supra* at 170; Memarzadeh & Berek, *supra*; Shridhar et al., *supra*. Staging is performed by reference to the classification system developed by the International Federation of Gynecology and Obstetrics. David H. Moore, Primary Surgical Management of Early Epithelial Ovarian Carcinoma, in Ovarian Cancer 203 (Stephen C. Rubin
20 & Gregory P. Sutton eds., 2d ed. 2001); Fleming et al. eds., *supra* at 188. Stage I ovarian cancer is characterized by tumor growth that is limited to the ovaries and is comprised of three substages. *Id.* In substage IA, tumor growth is limited to one ovary, there is no tumor on the external surface of the ovary, the ovarian capsule is intact, and no malignant cells are present
25 in ascites or peritoneal washings. *Id.* Substage IB is identical to A1, except that tumor growth is limited to both ovaries. *Id.* Substage IC refers to the presence of tumor growth limited to one or both ovaries, and
30

also includes one or more of the following characteristics: capsule rupture, tumor growth on the surface of one or both ovaries, and malignant cells present in ascites or peritoneal washings. Id.

Stage II ovarian cancer refers to tumor growth involving one or both ovaries, along with pelvic extension. Id. Substage IIA involves extension and/or implants on the uterus and/or fallopian tubes, with no malignant cells in the ascites or peritoneal washings, while substage IIB involves extension into other pelvic organs and tissues, again with no malignant cells in the ascites or peritoneal washings. Id. Substage IIC involves pelvic extension as in IIA or IIB, but with malignant cells in the ascites or peritoneal washings. Id.

Stage III ovarian cancer involves tumor growth in one or both ovaries, with peritoneal metastasis beyond the pelvis confirmed by microscope and/or metastasis in the regional lymph nodes. Id. Substage IIIA is characterized by microscopic peritoneal metastasis outside the pelvis, with substage IIIB involving macroscopic peritoneal metastasis outside the pelvis 2 cm or less in greatest dimension. Id. Substage IIIC is identical to IIIB, except that the metastasis is greater than 2 cm in greatest dimension and may include regional lymph node metastasis. Id. Lastly, Stage IV refers to the presence distant metastasis, excluding peritoneal metastasis. Id.

While surgical staging is currently the benchmark for assessing the management and treatment of ovarian cancer, it suffers from considerable drawbacks, including the invasiveness of the procedure, the potential for complications, as well as the potential for inaccuracy. Moore, *supra* at 206-208, 213. In view of these limitations, attention has turned to developing alternative staging methodologies through understanding differential gene expression in various stages of ovarian cancer and by obtaining various biomarkers to help better assess the progression of the disease. Vartiainen, J. et al., *Int'l J. Cancer*, 95(5): 313-16 (2001); Shridhar et al. *supra*; Baekelandt, M. et al., *J. Clin. Oncol.* 18(22): 3775-81.

The treatment of ovarian cancer typically involves a multiprong attack, with

surgical intervention serving as the foundation of treatment. Dennis S. Chi & William J. Hoskins, Primary Surgical Management of Advanced Epithelial Ovarian Cancer, in Ovarian Cancer 241 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). For example, in the case of epithelial ovarian cancer, which accounts for ~90% of cases of ovarian cancer, treatment typically consists of: (1) cytoreductive surgery, including total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy, and lymphadenectomy, followed by (2) adjuvant chemotherapy with paclitaxel and either cisplatin or carboplatin. Eltabbakh, G.H. & Awtrey, C.S., Expert Op. Pharmacother. 2(10): 109-24. Despite a clinical response rate of 80% to the adjuvant therapy, most patients experience tumor recurrence within three years of treatment. Id. Certain patients may undergo a second cytoreductive surgery and/or second-line chemotherapy. Memarzadeh & Berek, supra.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of ovarian cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop ovarian cancer, for diagnosing ovarian cancer, for monitoring the progression of the disease, for staging the ovarian cancer, for determining whether the ovarian cancer has metastasized, and for imaging the ovarian cancer. There is also a need for better treatment of ovarian cancer.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovaries; identify and monitor ovary tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered ovary tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to ovary cells and/or ovary tissue. These ovary specific nucleic acids (OSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the OSNA is genomic DNA, then the OSNA is an ovary specific gene (OSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to ovary. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 119 through 228. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 118. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding an OSP, or that selectively hybridize or exhibit substantial sequence similarity to an OSNA, as well as allelic variants of a nucleic acid molecule encoding an OSP, and allelic variants of an OSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes an OSP or that comprises a part of a nucleic acid sequence of an OSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of an OSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of an OSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic

acid molecule encodes all or a fragment of an OSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of an OSNA.

Another object of the invention is to provide methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly
5 produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is an OSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutin), fusion protein, homologous protein or a polypeptide encoded by
10 an allelic variant of an OSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid
15 molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and
20 treating ovarian cancer and non-cancerous disease states in ovaries. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring ovary tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered ovary tissue for treatment and research.

25 The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovaries. The invention provides methods of using the polypeptides of the invention to identify and/or monitor ovary tissue, and to produce engineered ovary tissue.

30 The agonists and antagonists of the instant invention may be used to treat ovarian cancer and non-cancerous disease states in ovaries and to produce engineered ovary tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

5

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular
10 terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed
15 according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor
20 Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology - 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using
25 Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and
30 techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in

the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 5 A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and
- 10 "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.
- 15 The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates,
- 20 phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially
- 25 duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.
- 30 A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may

comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

10 A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, e.g., ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

30 A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid

molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a
5 nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the
10 disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other
15 nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others.
20 Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, e.g. for use as probes or primers, or may be double-stranded, e.g. for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense
25 oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of
30 other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are

not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

10 The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotide linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See e.g., LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

25 Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

30 The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence.

In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, e.g., the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, e.g., for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

5
$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

10 In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C
15 would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

20 An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without
25 formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a
30 library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping

the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. See Sambrook *et al.*

- 5 (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. See also Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (see Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

- 15 As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

- 20 Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (e.g., for oligonucleotide probes) may be calculated by the formula: $T_m = 81.5^\circ\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/N)$, wherein N is change length and the $[\text{Na}^+]$ is 1 M or less. See Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. See, e.g., Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

30 The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μ g of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μ l of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μ g of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity

with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding an OSP or is an OSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR

reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. See, e.g., Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, e.g., Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, e.g., Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

5 The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination,
10 promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such
15 control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

20 The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of
25 vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and
30 thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression

vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that
5 serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding
10 generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in
15 its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

20 As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-
25 occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises an OSP encoded by a nucleic acid molecule of the instant
30 invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally

associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be

5 "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single

10 species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,

15 followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared

20 to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40

25 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such

30 modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antigens (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antigens which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences.

Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial

identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with
5 respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide
10 compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or
15 pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH₂NH--, --CH₂S--, --CH₂-CH₂--,
--CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂--, and --CH₂SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of
20 L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which
25 cyclize the peptide.

A "polypeptide mutant" or "muted" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A
30 muted protein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid

- sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.
- Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (e.g., a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino

-27-

acids may also be suitable components for polypeptides of the present invention.

Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate,

-N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine,

N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other

5 similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism
10 if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous,"
15 this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence
20 similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino
25 acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino
30 acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted

upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g., Pearson, Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are

5 conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 10 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is
 15 any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid
 20 substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutin thereof. *See, e.g., GCG Version 6.1.* Other programs include FASTA, discussed *supra*.

25 A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g., Altschul et al., J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al., Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

- 30 Expectation value: 10 (default)
 Filter: seg (default)
 Cost to open a gap: 11 (default)

-29-

- Cost to extend a gap: 1 (default)
Max. alignments: 100 (default)
Word size: 11 (default)
No. of descriptions: 100 (default)
5 Penalty Matrix: BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number
10 of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best
15 overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

20 An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv,
25 dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')₂ fragment is a bivalent fragment comprising two Fab
30 fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single

-30-

arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston *et al., Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al., Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that

purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (e.g., BSA) or a chemical such as polyethylene glycol (PEG).

5 A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist
10 of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than 1 μ M, preferably less than 100 nM and most preferably less than 10 nM.

15 The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

20 The term "ovary specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the ovary as compared to other tissues in the body. In a preferred embodiment, a "ovary specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "ovary specific" nucleic acid molecule or polypeptide is
25 expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels,
30 such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

5 One aspect of the invention provides isolated nucleic acid molecules that are specific to the ovary or to ovary cells or tissue or that are derived from such nucleic acid molecules. These isolated ovary specific nucleic acids (OSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid
10 molecule encodes a polypeptide that is specific to ovary, an ovary-specific polypeptide (OSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 119 through 228. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 118.

15 AN OSNA may be derived from a human or from another animal. In a preferred embodiment, the OSNA is derived from a human or other mammal. In a more preferred embodiment, the OSNA is derived from a human or other primate. In an even more preferred embodiment, the OSNA is derived from a human.

20 By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding an OSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode an OSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes an OSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that
25 selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 119 through 228. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 118.

30 In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under moderate stringency conditions. In a more preferred

embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 119 through 228. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 118. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding an OSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human OSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 119 through 228. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding an OSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 119 through 228, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding an OSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding an OSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to an OSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 118. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity

with an OSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 118, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with an OSNA, more preferably at least 95%,
5 more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with an OSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one
10 that exhibits sequence identity over its entire length to an OSNA or to a nucleic acid molecule encoding an OSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the OSNA or the nucleic acid molecule encoding an OSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at
15 least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 119 through 228 or demonstrates
20 significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 118. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the OSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated
25 species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other
30 species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is

experimentally produced by directed mutation of an OSNA. Further, the substantially similar nucleic acid molecule may or may not be an OSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is an OSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of an OSNA or a nucleic acid encoding an OSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes an OSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is an OSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 118. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is an OSP. However, in a preferred embodiment, the part encodes an OSP. In one aspect, the invention comprises a part of an OSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to an OSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of an OSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes an OSP. A part comprises at least 10 nucleotides, more preferably at least 15,

17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that
5 encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain
10 reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains
15 modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe,
20 the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be
25 limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment,
30 the labeled nucleic acid molecule may be used as a hybridization probe.

-37-

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as ^{-32}P -dATP, ^{-32}P -dCTP, ^{-32}P -dGTP, ^{-32}P -dTTP, ^{-32}P -3'dATP, ^{-32}P -ATP, ^{-32}P -CTP, ^{-32}P -GTP, ^{-32}P -UTP, ^{-35}S -dATP, α - ^{35}S -GTP, α - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3'

hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. See, e.g., Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. See Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.),

Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd

(1997); the disclosures of which are incorporated herein by reference in their entireties.

Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. See Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000),

5 the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including

10 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'.

Representative United States patents that teach the preparation of the above

15 phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by

20 reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or

25 more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and

30 methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S.

Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by $8\text{--}20^\circ\text{C}$ (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by $4\text{--}16^\circ\text{C}$ (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases

do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.,* Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably,

detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of an OSNA, such as deletions, insertions, translocations, and duplications of the OSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. See, e.g., Andreeff *et al.* (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, e.g., Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify OSNA in, and isolate OSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. See, e.g., Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to OSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in
5 their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding an OSP. In a more
10 preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 119 through 228. In another preferred embodiment, the probe or primer is derived from an OSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 118.

15 In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer
20 in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide
25 probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic
30 Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular

- Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

- The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

- Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention

can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides

encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones
5 *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000);
10 Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an
15 expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part
20 of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic
25 nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the
30 nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their

derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λ GT10 and λ GT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*,
5 typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic
10 genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will
15 typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2,
20 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*,
25 *leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors
30 are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-

transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

5 In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines
10 expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian
15 cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

20 Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

25 Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for
30 encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of

the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these
5 vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include
10 splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that
15 modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC
20 system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

25 Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast α -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the *CYC1* or *ADH1* gene.

30 Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early

gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the OSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, *ori*- or *ARS*-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to

be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or
5 identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, e.g., the gene III
10 protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See Barbas et al., Phage Display: A Laboratory Manual*, Cold Spring Harbor Laboratory Press (2001); *Kay et al. (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual*, Academic Press, Inc., (1996); *Abelson et al. (eds.), Combinatorial Chemistry* (Methods in Enzymology, Vol. 267) Academic Press (1996).
15 Vectors for yeast display, e.g. the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, e.g., the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet
20 derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring
25 proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic
30 fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See Li et al., J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from

those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. See Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by

5 reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g.* Cormack *et al.*, *Gene* 173: 33-38

10 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)).

15 Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.* Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified

20 GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. See also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of

25 each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application

30 Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide OSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database <http://www.abrf.org/ABRF/Research/Committees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and <http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html> (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one
5 may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is
10 capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

15 In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the
20 product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression
25 vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have
30 biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

- Transformation and other methods of introducing nucleic acids into a host cell
- 5 (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (See, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an
- 10 expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be
- 15 able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

- A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera*
- 20 *frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from
- 25 *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3
- 30 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from ovary are particularly preferred because they may provide a more native
5 post-translational processing. Particularly preferred are human ovary cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number
10 of texts and laboratory manuals in the art. *See, e.g.,* Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend
15 primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.,* Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

20 Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.,* with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent
25 strains are also available commercially (*e.g.,* Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse
30 treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols

(BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the
5 action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed
10 with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective
15 medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension
20 pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

25 Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated
30 transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA).

Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (<http://www.bio-rad.com/LifeScience/pdf/>

- 5 New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.,* Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA*
10 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.,* Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid
15 Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A
20 Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

- Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by
25 means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

- 30 Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is an ovary specific polypeptide (OSP). In an even more preferred embodiment, the

polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 119 through 228. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of an OSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 119 through 228. A polypeptide that comprises only a fragment of an entire OSP may or may not be a polypeptide that is also an OSP. For instance, a full-length polypeptide may be ovary-specific, while a fragment thereof may be found in other tissues as well as in ovary. A polypeptide that is not an OSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-OSP antibodies. However, in a preferred embodiment, the part or fragment is an OSP. Methods of determining whether a polypeptide is an OSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g., Lerner, Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for

-64-

the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies
5 (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

10 The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75
15 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, an OSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically
20 synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a
25 fragment of polypeptide of the invention, preferably an OSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably an OSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants,
30 fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be ovary-specific. In a preferred embodiment, the mutein is ovary-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 119 through 228. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is ovary-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, e.g., by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo*

mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.,* Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel
5 (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to an OSP. In an even more preferred embodiment, the
10 polypeptide is homologous to an OSP selected from the group having an amino acid sequence of SEQ ID NO: 119 through 228. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to an OSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 119 through 228. In an even
15 more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%,
20 more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to an OSP comprising an
25 amino acid sequence of SEQ ID NO: 119 through 228. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to an OSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that
30 hybridizes to an OSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the OSNA is selected from the group consisting of SEQ ID NO: 1 through 118. In another preferred

embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes an OSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the OSP is selected from the group consisting of SEQ ID NO: 119 through
5 228.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of
10 SEQ ID NO: 119 through 228. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the OSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster,
15 cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous
20 polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of an OSP. Further, the
25 homologous protein may or may not encode polypeptide that is an OSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is an OSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an
30 antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the

binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding an OSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 119 through 228. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 118.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an OSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 119 through 228, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983);

Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591,

BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

- 5 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS,
- 10 HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC,
- 15 SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

- The polypeptides, fragments, and fusion proteins of the present invention can be
- 20 conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

- The polypeptides, fragments, and fusion proteins of the present invention can also
- 25 usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-OSP antibodies.

- The polypeptides, fragments, and fusion proteins of the present invention can also
- 30 usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999),

incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

- 5 In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an OSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 119 through 228. In a preferred embodiment, the analog is one that comprises one or
- 10 more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to an OSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--,
- 15 --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of an OSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific
- 20 three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Kole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- 25 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A
- 30 Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and tBOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-Fmoc-L-glutamic acid or the corresponding tBOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated Fmoc synthesis of peptides using (Fmoc)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

- 2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

- Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

- The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is an OSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 119 through 228, or is a mutein, homologous

polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 118, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule

5 comprising a nucleic acid sequence of SEQ ID NO: 1 through 118.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino

10 acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and

15 usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety,

20 heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further

25 purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion

30 proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation

- of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.
- Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbri *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by

reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded
5 protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin
10 A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, myc, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin trpE, protein A, β -lactamase, α -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine
15 at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast α -mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. See, e.g., Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by
20 enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (e.g., a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind
25 the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the OSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize OSPs, their allelic
30 variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly OSPs, e.g. by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser

scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of OSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of OSPs.

- 5 One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon
10 linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-
15 102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. See, e.g., Scopes, Protein Purification, 2d ed. (1987).

- 20 Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, e.g., by HPLC.

- Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-
25 proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

- Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For
30 example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is an OSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 119 through 228, or a fragment, mutein, derivative, analog or fusion protein thereof.

10 The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on an OSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of an OSP may be indicative of cancer. Differential degradation of the C or N-terminus of an OSP may also be a marker or target for anticancer therapy. For example, an OSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies
15 may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-OSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human
25 ovary.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the

present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as
5 IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such
10 antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic
15 mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by
20 reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered
25 antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger
30 mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are

hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

5 As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

10 Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

20 Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular

advantage in detection of the proteins of the present invention, in human serum or tissues (Viking et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*,

4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujaime *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).

- 5 Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entirety.

Typically, phage-displayed antibody fragments are scFv fragments or Fab
10 fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

- 15 For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3): 157-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997); Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*,
20 *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entirety.

- Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.*, Li *et al.*, *Protein Expr. Purif.*
25 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entirety.

- Antibodies and fragments and derivatives thereof of the present invention can
30 also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavalondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,

Biotechnol. Appl. Biochem. 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entirety.

5 Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. See, e.g. Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entirety.

10 Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression
15 systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57
20 (1999), the disclosures of which are incorporated herein by reference in their entirety.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the
25 proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab')₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

30 It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated

nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

5 Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

10 Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature*
15 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2
20 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain
25 diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof,
30 whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions

including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate).

reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

10 As another example, when the antibodies of the present invention are used, e.g., for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

15 For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

20 Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568,
25 BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5,
30 Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H ,
5 and ^{125}I .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

10 As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the
15 application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, diphtheria
20 toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate,
25 and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more
30 of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to
5 paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in
10 prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind
15 specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present
20 invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody
25 molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human
30 organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding an OSP. In a preferred embodiment, the OSP comprises an amino

acid sequence selected from SEQ ID NO: 119 through 228, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise an OSNA of the invention, preferably an OSNA comprising a nucleotide sequence selected from the group
5 consisting of SEQ ID NO: 1 through 118, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human OSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-
10 human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g., Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, *Mouse Genetics and Transgenics: A Practical Approach*, Oxford University Press (2000); and Pinkert, *Transgenic Animal Technology: A Laboratory Handbook*, Academic Press (1999).
15

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g., Paterson et al., Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al., Biotechnology* 11: 1263-1270 (1993); Wright *et al., Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g., Van der Putten et al., Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g., Thompson et al., Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g., Lo, 1983, Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g., Ulmer et al., Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g., Lavitrano et al., Cell* 57: 717-723 (1989)).
20
25

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g., Campell et al., Nature* 380: 64-66 (1996); Wilmut *et al., Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e., a*
30

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene
5 may also be selectively introduced into and activated in a particular cell type by following, e.g., the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

10 Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using
15 techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

20 Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce ovaries of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels.
25 because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is
30 appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g., knockouts*) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC.

compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e.g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e.g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

10 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e.g., in the circulation, or intraperitoneally.

15 Alternatively, the cells can be incorporated into a matrix and implanted in the body, e.g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. See, e.g., U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

20 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

25 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 118 and SEQ ID NO: 119 through 228 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or
5 data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and
10 similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a
15 nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence
20 comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the
25 steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Ovarian Cancer

30 The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of an OSNA or an OSP in a human patient that has or may have ovarian cancer, or who is at risk of developing ovarian cancer, with the expression of an OSNA or an OSP in a normal human control. For purposes of the present invention, "expression of an OSNA" or "OSNA expression" means the quantity of OSG mRNA that
5 can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of an OSP" or "OSP expression" means the amount of OSP that can be measured by any method known in the art or the level of translation of an OSG OSNA that can be measured by any method known in the art.

10 The present invention provides methods for diagnosing ovarian cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of OSNA or OSP in cells, tissues, organs or bodily fluids compared with levels of OSNA or OSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of an OSNA or
15 OSP in the patient versus the normal human control is associated with the presence of ovarian cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in the structure of the mRNA of an OSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing,
20 alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in an OSP compared to an OSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the OSP or subcellular OSP localization.

25 In a preferred embodiment, the expression of an OSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 119 through 228, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the OSNA expression that is measured is the level of expression of an OSNA mRNA selected from SEQ ID NO: 1 through 118, or a
30 hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. OSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by

Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. See, e.g., Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. OSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of an OSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, e.g., aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, OSNA expression may be compared to a known control, such as normal ovary nucleic acid, to detect a change in expression.

10 In another preferred embodiment, the expression of an OSP is measured by determining the level of an OSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 119 through 228, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of OSNA or OSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of ovarian cancer. The expression level of an OSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the OSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. See, e.g., Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the OSP structure may be determined by any method known in the art, including, e.g., using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to an OSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-OSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the OSP will bind to the anti-OSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-OSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the OSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an OSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure OSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-OSP antibody is attached to a solid support and an allocated amount of a labeled OSP and a sample of interest are incubated with the solid support. The amount of labeled OSP detected which is attached to the solid support can be correlated to the quantity of an OSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of an OSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more OSNAs of interest. In this approach, all or a portion of one or more OSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of OSNA or OSP includes, without limitation, ovary tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, ovary cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary ovarian cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and breast. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. See Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in OSNAs or OSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

-100-

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of an OSNA or OSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other OSNA or OSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular OSNA or OSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a patient suspected of having ovarian cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP and then ascertaining whether the patient has ovarian cancer from the expression level of the OSNA or OSP. In general, if high expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether ovarian cancer has metastasized in a patient. One may identify whether the ovarian cancer has metastasized by measuring the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a variety of tissues. The presence of an OSNA or OSP in a

-101-

certain tissue at levels higher than that of corresponding noncancerous tissue (e.g., the same tissue from another individual) is indicative of metastasis if high level expression of an OSNA or OSP is associated with ovarian cancer. Similarly, the presence of an OSNA or OSP in a tissue at levels lower than that of corresponding noncancerous tissue is
5 indicative of metastasis if low level expression of an OSNA or OSP is associated with ovarian cancer. Further, the presence of a structurally altered OSNA or OSP that is associated with ovarian cancer is also indicative of metastasis.

In general, if high expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of
10 expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at
15 least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The OSNA or OSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with ovarian cancers or
20 other ovary related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of ovarian disorders.

Staging

25 The invention also provides a method of staging ovarian cancer in a human patient. The method comprises identifying a human patient having ovarian cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more OSNAs or OSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and
30 the expression level of one or more OSNAs or OSPs is determined for each stage to obtain a standard expression level for each OSNA and OSP. Then, the OSNA or OSP expression levels are determined in a biological sample from a patient whose stage of

-102-

cancer is not known. The OSNA or OSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the OSNAs and OSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations
5 of an OSNA or OSP to determine the stage of an ovarian cancer.

Monitoring

Further provided is a method of monitoring ovarian cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient
10 to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the ovarian cancer. The method comprises identifying a human patient that one wants to monitor for ovarian cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for
15 expression levels of one or more OSNAs or OSPs, and comparing the OSNA or OSP levels over time to those OSNA or OSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in an OSNA or OSP that are associated with ovarian cancer.

If increased expression of an OSNA or OSP is associated with metastasis,
20 treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or
25 failure to progress to a neoplastic lesion. If decreased expression of an OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of OSNAs or OSPs are
30 determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of ovarian cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of an OSNA and/or OSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more OSNAs and/or OSPs are detected. The presence of higher (or lower) OSNA or OSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly ovarian cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more OSNAs and/or OSPs of the invention can also be monitored by analyzing levels of expression of the OSNAs and/or OSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in an OSG, thereby determining if a human with the genetic lesion is susceptible to developing ovarian cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing ovarian cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the OSGs of this invention, a chromosomal rearrangement of OSG, an aberrant modification of OSG (such as of the methylation pattern of the genomic DNA), or allelic loss of an OSG. Methods to detect such lesions in the OSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

25 Methods of Detecting Noncancerous Ovarian Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a patient suspected of having or known to have a noncancerous ovarian disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP, comparing the expression level or structural alteration of the OSNA or OSP to a normal

ovary control, and then ascertaining whether the patient has a noncancerous ovarian disease. In general, if high expression relative to a control of an OSNA or OSP is indicative of a particular noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and
5 more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of a noncancerous ovarian disease, a diagnostic assay is considered positive if
10 the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether an OSNA and/or OSP
15 is associated with a particular noncancerous ovarian disease by obtaining ovary tissue from a patient having a noncancerous ovarian disease of interest and determining which OSNAs and/or OSPs are expressed in the tissue at either a higher or a lower level than in normal ovary tissue. In another embodiment, one may determine whether an OSNA or OSP exhibits structural alterations in a particular noncancerous ovarian disease state by
20 obtaining ovary tissue from a patient having a noncancerous ovarian disease of interest and determining the structural alterations in one or more OSNAs and/or OSPs relative to normal ovary tissue.

Methods for Identifying Ovary Tissue

25 In another aspect, the invention provides methods for identifying ovary tissue. These methods are particularly useful in, e.g., forensic science, ovary cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a
30 sample is ovary tissue or has ovary tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising ovary tissue or having ovary tissue-like characteristics, determining whether the sample expresses one or more OSNAs and/or OSPs, and, if the sample expresses one or more OSNAs and/or OSPs, concluding

that the sample comprises ovary tissue. In a preferred embodiment, the OSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 119 through 228, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 118, or a
5 hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses an OSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether an OSP is expressed. Determining whether a
10 sample expresses an OSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the OSP has an amino acid sequence selected from SEQ ID NO: 119 through 228, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two OSNAs and/or OSPs is determined. In a
15 more preferred embodiment, the expression of at least three, more preferably four and even more preferably five OSNAs and/or OSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is ovary tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are
20 recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into ovary tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, e.g., in producing new ovary tissue by tissue engineering. These agents include, e.g., growth and differentiation factors, extracellular matrix proteins and culture medium.
25 Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Ovary Tissue

30 In another aspect, the invention provides methods for producing engineered ovary tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing an OSNA or an OSG into the cells, and growing the cells under conditions in

which they exhibit one or more properties of ovary tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal ovary tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered ovary tissue or cells comprises one of these cell types. In another
5 embodiment, the engineered ovary tissue or cells comprises more than one ovary cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the ovary cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more OSPs are introduced into cells,
10 preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode OSPs having amino acid sequences selected from SEQ ID NO: 119 through 228, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 118, or hybridizing nucleic acids, allelic variants or parts
15 thereof. In another highly preferred embodiment, an OSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial ovary tissue may be used to treat patients who have lost some or all of their ovary function.

20 Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a
25 preferred embodiment, the pharmaceutical composition comprises an OSNA or part thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 118, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises an
30 OSP or fragment thereof. In a more preferred embodiment, the OSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 119 through 228, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of

the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-OSP antibody, preferably an antibody that specifically binds to an OSP having an amino acid that is selected from the group consisting of SEQ ID NO: 119 through 228, or an antibody that binds to a polypeptide
5 that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

10 Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed.
15 (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and
20 parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

25 Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and
30 tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

- 5 Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (PovidoneTM), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

- 10 Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

- 15 Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

- 25 Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

- Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.
- 30

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

10 Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil
15 base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene
20 glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of
25 the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature
30 of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot

injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

5 For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration
10 of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

15 For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as
20 cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for
25 treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the
30 present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts

tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

- 5 The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

- A "therapeutically effective dose" refers to that amount of active ingredient, for example OSP polypeptide, fusion protein, or fragments thereof, antibodies specific for
10 OSP, agonists, antagonists or inhibitors of OSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

- The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed
15 by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

- For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in
20 one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

- The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of
25 circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

- 30 The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age,

weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

- 5 Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be
10 administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,
15 conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

20 Therapeutic Methods

- The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of ovary function. As used herein,
25 "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

- 30 The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication

incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further
5 described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g., Doronin et al., J. Virol.* 75: 3314-24 (2001).

10 In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of an OSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of an OSP are
15 administered, for example, to complement a deficiency in the native OSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g., Cid-Arregui, supra.* In a preferred embodiment, the nucleic acid molecule encodes an OSP having the amino acid sequence of SEQ ID NO: 119 through
20 228, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express an OSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement
25 defects in OSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode an OSP having the amino acid sequence of SEQ ID NO: 119 through 228, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of an OSG
30 antisense nucleic acid, are administered to downregulate transcription and/or translation of an OSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of an OSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

- 5 Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to OSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are
10 incorporated herein by reference in their entireties.

- Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the OSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9
15 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

- In a preferred embodiment, the antisense molecule is derived from a nucleic acid
20 molecule encoding an OSP, preferably an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 118, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

25 *Polypeptide Administration*

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an OSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant OSP defect.

- 30 Protein compositions are administered, for example, to complement a deficiency in native OSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to OSP. The immune

response can be used to modulate activity of OSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate OSP.

5 In a preferred embodiment, the polypeptide is an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 118, or a part, allelic variant, substantially similar or hybridizing nucleic acid
10 thereof.

Antibody, Agonist and Antagonist Administration

 In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is
15 administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of OSP, or to target therapeutic agents to sites of OSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred
20 embodiment, the antibody specifically binds to an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 118, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

 The present invention also provides methods for identifying modulators which bind to an OSP or have a modulatory effect on the expression or activity of an OSP.
25 Modulators which decrease the expression or activity of OSP (antagonists) are believed to be useful in treating ovarian cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of an OSP can also be designed, synthesized and tested for use in the imaging and treatment of ovarian
30 cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the OSPs identified herein. Molecules identified in the library as being capable of binding to an OSP are key candidates for

further evaluation for use in the treatment of ovarian cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of an OSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of OSP is administered. Antagonists of OSP can be produced using methods generally known in the art. In particular, purified OSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of an OSP.

In other embodiments a pharmaceutical composition comprising an agonist of an OSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP comprising an amino acid sequence of SEQ ID NO: 119 through 228, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 118, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Ovary Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the ovary or to specific cells in the ovary. In a preferred embodiment, an anti-OSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if ovary tissue needs to be selectively destroyed. This would be useful for targeting and killing ovarian cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting ovary cell function.

In another embodiment, an anti-OSP antibody may be linked to an imaging agent that can be detected using, e.g., magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring ovary function, identifying ovarian cancer tumors, and identifying noncancerous ovarian diseases.

EXAMPLES

Example 1: Gene Expression analysis

OSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the OSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), and differential expression in cancer tissue (CLASP 5). cDNA libraries were divided into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within

-118-

each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

- (a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.
- (b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.
- (c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

CLASP Expression percentage levels for DEX0277 genes

DEX0277_111	SEQ ID NO: 111	OVR .001	LIV .0011	TST .0011	UNC .0011
DEX0277_53	SEQ ID NO: 53	OVR .001	LIV .0011	TST .0011	UNC .0011
DEX0277_79	SEQ ID NO: 79	BRN .0004	SPL .0021	LMN .0028	SYN .0028

Abbreviation for tissues:

BLO Blood; BRN Brain; CON Connective Tissue; CRD Heart; FTS Fetus; INL Intestine, Large; INS Intestine, Small; KID Kidney; LIV Liver; LNG Lung; MAM Breast; MSL Muscles; NRV Nervous Tissue; OVR Ovary; PRO Prostate; STO Stomach; THR Thyroid Gland; TNS Tonsil / Adenoids; UTR Uterus

5

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the OSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the OSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for
5 that same tissue from the same individual.

In the analysis of matching samples, the OSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal
10 adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 118 being a diagnostic
15 marker for cancer.

Example 3: Protein Expression

The OSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the OSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the OSNA coding sequence, codons for two amino acids, Met-Ala, flanking
20 the NH₂-terminus of the coding sequence of OSNA, and six histidines, flanking the COOH-terminus of the coding sequence of OSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is
25 confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of OSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column
30 volumes of wash buffer. OSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.,* WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes

-122-

use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the

5 hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and

10 Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

Based on the nucleotide sequences found by mRNA subtractions the following extended nucleic acid sequences and amino acid sequences were determined.

	SEQ ID NO0277_1	SEQ ID NO0125_1	SEQ ID NO0277_119
15	SEQ ID NO0277_2	SEQ ID NO0125_2	SEQ ID NO0277_120
	SEQ ID NO0277_3	SEQ ID NO0125_3	SEQ ID NO0277_121
	SEQ ID NO0277_4	SEQ ID NO0125_4	SEQ ID NO0277_122
	SEQ ID NO0277_5	flex SEQ ID NO0125_4	
	SEQ ID NO0277_6	SEQ ID NO0125_5	SEQ ID NO0277_123
20	SEQ ID NO0277_7	SEQ ID NO0125_6	SEQ ID NO0277_124
	SEQ ID NO0277_8	SEQ ID NO0125_7	SEQ ID NO0277_125
	SEQ ID NO0277_9	SEQ ID NO0125_8	SEQ ID NO0277_126
	SEQ ID NO0277_10	SEQ ID NO0125_9	SEQ ID NO0277_127
	SEQ ID NO0277_11	SEQ ID NO0125_10	SEQ ID NO0277_128
25	SEQ ID NO0277_12	SEQ ID NO0125_11	SEQ ID NO0277_129
	SEQ ID NO0277_13	flex SEQ ID NO0125_11	
	SEQ ID NO0277_14	SEQ ID NO0125_12	SEQ ID NO0277_130
	SEQ ID NO0277_15	SEQ ID NO0125_13	
	SEQ ID NO0277_16	SEQ ID NO0125_14	SEQ ID NO0277_131
30	SEQ ID NO0277_17	SEQ ID NO0125_15	SEQ ID NO0277_132
	SEQ ID NO0277_18	SEQ ID NO0125_16	
	SEQ ID NO0277_19	SEQ ID NO0125_17	SEQ ID NO0277_133

-123-

SEQ ID NO0277_20 SEQ ID NO0125_18 SEQ ID NO0277_134
SEQ ID NO0277_21 flex SEQ ID NO0125_18 SEQ ID NO0277_135
SEQ ID NO0277_22 SEQ ID NO0125_19 SEQ ID NO0277_136
SEQ ID NO0277_23 SEQ ID NO0125_20 SEQ ID NO0277_137
5 SEQ ID NO0277_24 SEQ ID NO0125_21 SEQ ID NO0277_138
SEQ ID NO0277_25 SEQ ID NO0125_22
SEQ ID NO0277_26 flex SEQ ID NO0125_22
SEQ ID NO0277_27 SEQ ID NO0125_23
SEQ ID NO0277_28 SEQ ID NO0125_24 SEQ ID NO0277_139
10 SEQ ID NO0277_29 flex SEQ ID NO0125_24 SEQ ID NO0277_140
SEQ ID NO0277_30 SEQ ID NO0125_25 SEQ ID NO0277_141
SEQ ID NO0277_31 SEQ ID NO0125_26
SEQ ID NO0277_32 SEQ ID NO0125_27 SEQ ID NO0277_142
SEQ ID NO0277_33 flex SEQ ID NO0125_27
15 SEQ ID NO0277_34 SEQ ID NO0125_28 SEQ ID NO0277_143
SEQ ID NO0277_35 SEQ ID NO0125_29 SEQ ID NO0277_144
SEQ ID NO0277_36 SEQ ID NO0125_30 SEQ ID NO0277_145
SEQ ID NO0277_37 SEQ ID NO0125_31 SEQ ID NO0277_146
SEQ ID NO0277_38 SEQ ID NO0125_32 SEQ ID NO0277_147
20 SEQ ID NO0277_39 SEQ ID NO0125_33 SEQ ID NO0277_148
SEQ ID NO0277_40 SEQ ID NO0125_34 SEQ ID NO0277_149
SEQ ID NO0277_41 SEQ ID NO0125_35 SEQ ID NO0277_150
SEQ ID NO0277_42 SEQ ID NO0125_36 SEQ ID NO0277_151
SEQ ID NO0277_43 SEQ ID NO0125_37 SEQ ID NO0277_152
25 SEQ ID NO0277_44 SEQ ID NO0125_38 SEQ ID NO0277_153
SEQ ID NO0277_45 SEQ ID NO0125_39 SEQ ID NO0277_155
SEQ ID NO0277_46 SEQ ID NO0125_40
SEQ ID NO0277_47 SEQ ID NO0125_41 SEQ ID NO0277_156
SEQ ID NO0277_48 SEQ ID NO0125_42 SEQ ID NO0277_157
30 SEQ ID NO0277_49 SEQ ID NO0125_43 SEQ ID NO0277_158
SEQ ID NO0277_50 SEQ ID NO0125_44 SEQ ID NO0277_160
SEQ ID NO0277_51 SEQ ID NO0125_45 SEQ ID NO0277_162

-124-

- SEQ ID NO0277_52 SEQ ID NO0125_46 SEQ ID NO0277_163
SEQ ID NO0277_53 flex SEQ ID NO0125_46 SEQ ID NO0277_164
SEQ ID NO0277_54 SEQ ID NO0125_47 SEQ ID NO0277_165
SEQ ID NO0277_55 SEQ ID NO0125_48 SEQ ID NO0277_166
5 SEQ ID NO0277_56 SEQ ID NO0125_49
SEQ ID NO0277_57 SEQ ID NO0125_50 SEQ ID NO0277_167
SEQ ID NO0277_58 flex SEQ ID NO0125_50 SEQ ID NO0277_168
SEQ ID NO0277_59 SEQ ID NO0125_51
SEQ ID NO0277_60 SEQ ID NO0125_52
10 SEQ ID NO0277_61 SEQ ID NO0125_53 SEQ ID NO0277_169
SEQ ID NO0277_62 SEQ ID NO0125_54
SEQ ID NO0277_63 SEQ ID NO0125_55 SEQ ID NO0277_170
SEQ ID NO0277_64 flex SEQ ID NO0125_55 SEQ ID NO0277_171
SEQ ID NO0277_65 SEQ ID NO0125_56 SEQ ID NO0277_172
15 SEQ ID NO0277_66 SEQ ID NO0125_57 SEQ ID NO0277_173
SEQ ID NO0277_67 SEQ ID NO0125_58
SEQ ID NO0277_68 SEQ ID NO0125_59 SEQ ID NO0277_174
SEQ ID NO0277_69 SEQ ID NO0125_60 SEQ ID NO0277_175
SEQ ID NO0277_70 SEQ ID NO0125_61 SEQ ID NO0277_176
20 SEQ ID NO0277_71 SEQ ID NO0125_62 SEQ ID NO0277_177
SEQ ID NO0277_72 SEQ ID NO0125_63 SEQ ID NO0277_178
SEQ ID NO0277_73 SEQ ID NO0125_64
SEQ ID NO0277_74 SEQ ID NO0125_65 SEQ ID NO0277_181
SEQ ID NO0277_75 SEQ ID NO0125_66 SEQ ID NO0277_183
25 SEQ ID NO0277_76 flex SEQ ID NO0125_66 SEQ ID NO0277_184
SEQ ID NO0277_77 SEQ ID NO0125_67 SEQ ID NO0277_185
SEQ ID NO0277_78 SEQ ID NO0125_68 SEQ ID NO0277_186
SEQ ID NO0277_79 flex SEQ ID NO0125_68 SEQ ID NO0277_187
SEQ ID NO0277_80 SEQ ID NO0125_69
30 SEQ ID NO0277_81 SEQ ID NO0125_70 SEQ ID NO0277_188
SEQ ID NO0277_82 flex SEQ ID NO0125_70 SEQ ID NO0277_189
SEQ ID NO0277_83 SEQ ID NO0125_71 SEQ ID NO0277_190

-125-

- SEQ ID NO0277_84 SEQ ID NO0125_72 SEQ ID NO0277_191
 SEQ ID NO0277_85 SEQ ID NO0125_73 SEQ ID NO0277_192
 SEQ ID NO0277_86 SEQ ID NO0125_74 SEQ ID NO0277_193
 SEQ ID NO0277_87 SEQ ID NO0125_75 SEQ ID NO0277_194
 5 SEQ ID NO0277_88 SEQ ID NO0125_76 SEQ ID NO0277_196
 SEQ ID NO0277_89 SEQ ID NO0125_77 SEQ ID NO0277_197
 SEQ ID NO0277_90 flex SEQ ID NO0125_77 SEQ ID NO0277_198
 SEQ ID NO0277_91 SEQ ID NO0125_78 SEQ ID NO0277_199
 SEQ ID NO0277_92 flex SEQ ID NO0125_78
 10 SEQ ID NO0277_93 SEQ ID NO0125_79
 SEQ ID NO0277_94 SEQ ID NO0125_80 SEQ ID NO0277_201
 SEQ ID NO0277_95 SEQ ID NO0125_81 SEQ ID NO0277_202
 SEQ ID NO0277_96 flex SEQ ID NO0125_81 SEQ ID NO0277_203
 SEQ ID NO0277_97 SEQ ID NO0125_82
 15 SEQ ID NO0277_98 flex SEQ ID NO0125_82 SEQ ID NO0277_204
 SEQ ID NO0277_99 SEQ ID NO0125_83 SEQ ID NO0277_205
 SEQ ID NO0277_100 flex SEQ ID NO0125_83 SEQ ID NO0277_206
 SEQ ID NO0277_101 SEQ ID NO0125_84 SEQ ID NO0277_207
 SEQ ID NO0277_102 SEQ ID NO0125_85
 20 SEQ ID NO0277_103 SEQ ID NO0125_86 SEQ ID NO0277_209
 SEQ ID NO0277_104 SEQ ID NO0125_87 SEQ ID NO0277_211
 SEQ ID NO0277_105 SEQ ID NO0125_88 SEQ ID NO0277_212
 SEQ ID NO0277_106 SEQ ID NO0125_89 SEQ ID NO0277_213
 SEQ ID NO0277_107 SEQ ID NO0125_90 SEQ ID NO0277_215
 25 SEQ ID NO0277_108 SEQ ID NO0125_91 SEQ ID NO0277_216
 SEQ ID NO0277_109 SEQ ID NO0125_92 SEQ ID NO0277_217
 SEQ ID NO0277_110 SEQ ID NO0125_93 SEQ ID NO0277_218
 SEQ ID NO0277_111 flex SEQ ID NO0125_93 SEQ ID NO0277_219
 SEQ ID NO0277_112 SEQ ID NO0125_94 SEQ ID NO0277_220
 30 SEQ ID NO0277_113 flex SEQ ID NO0125_94 SEQ ID NO0277_221
 SEQ ID NO0277_114 SEQ ID NO0125_95 SEQ ID NO0277_222
 SEQ ID NO0277_115 SEQ ID NO0125_96 SEQ ID NO0277_224

-126-

SEQ ID NO0277_116 SEQ ID NO0125_97 SEQ ID NO0277_226
 SEQ ID NO0277_117 SEQ ID NO0125_98 SEQ ID NO0277_227
 SEQ ID NO0277_118 SEQ ID NO0125_99 SEQ ID NO0277_228

5 The following Jamison-Wolf antigenic sites were also determined.

	DEX0277_121	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	79-112	1.07 34
	115-179	1.03 65
10	DEX0277_131	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	22-32	1.10 11
	DEX0277_143	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
15	39-52	1.22 14
	DEX0277_144	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	7-28	1.04 22
	DEX0277_147	Antigenicity Index(Jameson-Wolf)
20	positions	AI avg length
	19-31	1.08 13
	37-48	1.07 12
	DEX0277_148	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
25	57-78	1.06 22
	DEX0277_149	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	2-15	1.12 14
	DEX0277_150	Antigenicity Index(Jameson-Wolf)
30	positions	AI avg length
	3-16	1.13 14
	DEX0277_153	Antigenicity Index(Jameson-Wolf)

-127-

	positions	AI avg length
	4-21	1.03 18
	DEX0277_154	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
5	27-37	1.12 11
	DEX0277_155	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	19-43	1.10 25
	61-72	1.02 12
10	DEX0277_159	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	23-38	1.05 16
	DEX0277_160	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
15	56-68	1.05 13
	DEX0277_161	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	60-70	1.01 11
	DEX0277_163	Antigenicity Index(Jameson-Wolf)
20	positions	AI avg length
	15-24	1.19 10
	DEX0277_164	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	60-71	1.09 12
25	DEX0277_166	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	66-77	1.23 12
	37-61	1.10 25
	DEX0277_168	Antigenicity Index(Jameson-Wolf)
30	positions	AI avg length
	126-142	1.13 17
	456-468	1.05 13

-128-

	DEX0277_172	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	43-63	1.28 21
	25-38	1.26 14
5	5-20	1.12 16
	DEX0277_176	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	32-51	1.15 20
	DEX0277_179	Antigenicity Index(Jameson-Wolf)
10	positions	AI avg length
	105-135	1.07 31
	DEX0277_181	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	59-73	1.17 15
15	DEX0277_183	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	49-63	1.03 15
	DEX0277_184	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
20	54-91	1.11 38
	DEX0277_187	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	23-53	1.08 31
	DEX0277_188	Antigenicity Index(Jameson-Wolf)
25	positions	AI avg length
	15-44	1.14 30
	DEX0277_189	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	308-320	1.26 13
30	674-696	1.08 23
	63-78	1.08 16
	254-266	1.07 13

-129-

	441-451	1.07	11
	707-728	1.01	22
	DEX0277_190	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
5	4-26	1.03	23
	DEX0277_193	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	26-51	1.26	26
	DEX0277_194	Antigenicity Index(Jameson-Wolf)	
10	positions	AI avg length	
	14-27	1.24	14
	DEX0277_197	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	81-92	1.03	12
15	15-36	1.02	22
	DEX0277_198	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	39-52	1.15	14
	DEX0277_202	Antigenicity Index(Jameson-Wolf)	
20	positions	AI avg length	
	25-49	1.05	25
	DEX0277_204	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	35-73	1.21	39
25	169-193	1.14	25
	91-107	1.09	17
	114-149	1.04	36
	DEX0277_206	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
30	13-22	1.14	10
	DEX0277_207	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	

-130-

	15-44	1.14	30
	61-89	1.06	29
	116-130	1.06	15
	DEX0277_215	Antigenicity Index(Jameson-Wolf)	
5	positions	AI avg length	
	23-33	1.30	11
	61-86	1.10	26
	DEX0277_216	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
10	62-75	1.10	14
	DEX0277_218	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	15-24	1.19	10
	DEX0277_219	Antigenicity Index(Jameson-Wolf)	
15	positions	AI avg length	
	60-71	1.09	12
	DEX0277_220	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	2-16	1.04	15
20	DEX0277_221	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	353-366	1.15	14
	67-85	1.01	19
	DEX0277_222	Antigenicity Index(Jameson-Wolf)	
25	positions	AI avg length	
	27-73	1.11	47
	DEX0277_224	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	14-27	1.24	14
30	DEX0277_228	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg length	
	3-29	1.04	27

-131-

In addition, the following helical regions were predicted.

	DEX0277_123	PredHel=2	Topology=o26-48i55-74o
	DEX0277_132	PredHel=1	Topology=o10-27i
	DEX0277_140	PredHel=7	Topology=o37-59i72-94o120-142i149-171o205-
5	227i240-262o282-304i		
	DEX0277_145	PredHel=2	Topology=o5-27i75-97o
	DEX0277_148	PredHel=1	Topology=o10-29i
	DEX0277_156	PredHel=3	Topology=o4-23i36-55o59-78i
	DEX0277_157	PredHel=4	Topology=i13-35o55-77i79-101o116-138i
10	DEX0277_160	PredHel=1	Topology=i7-29o
	DEX0277_161	PredHel=1	Topology=o5-23i
	DEX0277_164	PredHel=1	Topology=o15-37i
	DEX0277_168	PredHel=2	Topology=i274-296o411-433i
	DEX0277_170	PredHel=1	Topology=i13-35o
15	DEX0277_186	PredHel=1	Topology=o10-29i
	DEX0277_190	PredHel=1	Topology=i30-52o
	DEX0277_192	PredHel=1	Topology=i7-24o
	DEX0277_196	PredHel=1	Topology=i45-67o
	DEX0277_199	PredHel=3	Topology=i2-24o28-45i52-74o
20	DEX0277_213	PredHel=3	Topology=i44-66o81-103i105-127o
	DEX0277_219	PredHel=1	Topology=o15-37i

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

- 25 RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 118. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds;
- 30 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then
5 cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are
10 nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium
15 iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical
20 Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological 25 Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 µg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells
30 are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results.

-133-

The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μ l of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μ l of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 μ g/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μ g/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally,

-134-

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of
5 administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-
10 release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-
15 release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and
20 EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is
25 formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and
30 other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is

shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as
5 liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic
10 acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such
15 as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or
20 stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial
25 having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture
30 is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 $\mu\text{g/kg}$ of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and

-137-

separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove

detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

10 **Example 12: Method of Treatment Using Gene Therapy-*In Vivo***

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

15 The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) *Cardiovasc. Res.* 35 (3): 470-479, Chao J et al. (1997) *Pharmacol. Res.* 35 (6): 517-522, Wolff J. A. (1997) *Neuromuscul. Disord.* 7 (5): 314-318, Schwartz B. et al. (1996) *Gene Ther.* 3 (5): 405-411, Tsurumi Y. et al. (1996) *Circulation* 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et

al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 $\mu\text{g/kg}$ body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be

-140-

determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

30 Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea

pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

5 Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver et al., *Biotechnology (NY)* 11: 1263-1270 (1993); Wright et al., *Biotechnology (NY)* 9: 830-
10 834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, *Mol Cell Biol.* 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a
15 gene gun (see, e. g., Ulmer et al., *Science* 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., *Cell* 57: 717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," *Intl. Rev. Cytol.* 115: 171-229 (1989), which is incorporated by reference herein in its
20 entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., *Nature* 380: 64-66 (1996); Wilmut et al., *Nature* 385: 810813 (1997)).

25 The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a
30 particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of

interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed
5 for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences
10 required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to
15 verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated
20 immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than
25 one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA
30 analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or

alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, 5 YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be 10 introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & 15 Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the 20 cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function 25 of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually 30 and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments,

-145-

which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes
5 an amino acid sequence of SEQ ID NO: 119 through 228;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID
NO: 1 through 118;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid
molecule of (a) or (b); or
 - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is a cDNA.
- 15 3. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of an ovary specific nucleic acid
(OSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule according to claim 1
under conditions in which the nucleic acid molecule will selectively hybridize to an
ovary specific nucleic acid; and
 - 30 (b) detecting hybridization of the nucleic acid molecule to an OSNA in the
sample, wherein the detection of the hybridization indicates the presence of an OSNA in
the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

5

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60%
15 sequence identity to of SEQ ID NO: 119 through 228; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 118.

12. An antibody or fragment thereof that specifically binds to the polypeptide
20 according to claim 11.

13. A method for determining the presence of an ovary specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under
25 conditions in which the antibody will selectively bind to the ovary specific protein; and

(b) detecting binding of the antibody to an ovary specific protein in the sample, wherein the detection of binding indicates the presence of an ovary specific protein in the sample.

30 14. A method for diagnosing and monitoring the presence and metastases of ovarian cancer in a patient, comprising the steps of:

-148-

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the ovary specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of ovarian cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with ovarian cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the ovarian cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

20

SEQUENCE LISTING

<110> Salceda, Susana
 Macina, Roberto
 Recipon, Heirve
 Cafferkey, Robert
 Sun, Yongming
 Liu, Chenghua
 diaDexus, Inc.

<120> Compositions and Methods Relating to Ovarian Specific Genes and Proteins

<130> DEX-0278

<150> 60/249,997

<151> 2000-11-20

<160> 228

<170> PatentIn version 3.1

<210> 1

<211> 811

<212> DNA

<213> Homo sapien

<400> 1

```

gaagagaagg gaagagaatg aagatactat ggggacctgg ttatctagat gatgctcgag      60
cggcgcagtg tgatggatag cgcccgcccg ggcagggttac cacttcactc caggcctggg      120
tgacaatatc tgagactctg tctcaaaaaa aaaaaaaaca actagttttt taaattttat      180
attttttaga acttttattc ccttaagtct caaacctcct tagggccctt ggtggtcagg      240
ttgtagattg tgagttactc cctttattct cgaaaagtgc tgtgggaacc gagaattgtg      300
gttctccact ttttgtggat tttttcttca aattgtggaa gaatgtgcct caactcgtgg      360
ttctcaactg gcattattat gtggaaatth tagaagctgg agaagaaaca ctttatatat      420
tacctcgtgg tcttaaacca ttcttcattt ttctaggggt agcaatattt aagttccaga      480
aaaaactcac cgtgggttaa agtcataaa ttgggtaatt agtactcaag aacgtgacag      540
accgtggggt tctttaaagg tgataataga gtgtacattt cattgggaat ttacattcat      600
acgtgcacgt gctgttatgc tgcattatth gggggaataa aagatttata atcttggggc      660
tggggcgcag tggggctcat gccttgtaa ttccagcac tttggggagg gctggaggca      720
ggtgggtatc acaaagtca gcagttccga gaccagtctg ggccaacctt ggggtggagac      780
cccgtttcta ctaaaaacta caagaacgtg g                                811

```

<210> 2

<211> 222

<212> DNA
 <213> Homo sapien

<400> 2
 gcgagggtacc ctgattcaaa tataattcca tgagaagctg gactaaggac atatattcat 60
 tcattcaata ttcattgtgtg tgtgtgttag agacaggggc ttgctctgtc ggccagattg 120
 gtttagatgt gtcacccctg atagatcaat gaaacgccgt atccgagggg gcctggggta 180
 aaaaaagggg ggaaaacaag gacagagagc agaagggggg ga 222

<210> 3
 <211> 1659
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (465)..(551)
 <223> a, c, g or t

<400> 3
 tgaagtcaagt tcatatccag ttcctcaacc agatgaccct tacattagca gaccctccct 60
 tcaagctggc tgactaacca ggcacttcaa gcaaccttca accacggaag ctccaaccac 120
 gccctaccaa cgcaacacag cttacgcaac ctgcactgcg aacacgctgc cggtgacgc 180
 acgcacctac tagcctacac gcctagcacc ttgcacgccc tacagcacgc caaacggcac 240
 gcagcaccta cgcacaccg caccctgcct ccacgcctct ggctctctg cgccacctag 300
 cgctgcgcc tccgagctcc cctcctgca cctgccttcc cctttatcca tacctgcgcc 360
 cagccctcc cttacgcca caccacctc acactatcac ccccatcat cctcgcaatc 420
 accacaacgc cccttggtgc gcccgcct tacgacgcct cccannnnnn nnnnnnnnnn 480
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
 nnnnnnnnnn ntctaaaacc ttccaccga catcccttac caaacctatc aacatttact 600
 tccccctaa tgcgctatcc ccaataattc caaactaacc ctgccccac tacgcccccc 660
 caaatctcc ccccgaccac cacacgcaac atcctcgtc accctactcc caccacctac 720
 acaacactca acactctcta ctacctctg accctcacac tccctcacc tctcttctc 780
 cactcattca cccccccatt acaactccac caacacactc gtcacacctc acccccatcc 840
 acaacgacct ccaccaccac cccactccca ccaactcatc agctacccc tctcctctc 900
 taccctacct acactaccac tccccccacc ctccctccc cacccttac tctcctccac 960
 ccatctctct ctactactta cctcaccact tctctcatc cccctcctc acgctcccc 1020

3

```

actctctctc tcctccgccc cacctcctcc ctacgcccc a ctcctcacct tccctctctc 1080
ttctatctac ctccgctott cccatgatca ccacatccac tcacatatac ccactcacta 1140
cacttctcac ggactctcgc actctcctca tccgtctcta catcctcaat cgcttcaccc 1200
cagcttcaca ctctatccaa cctacacaac tgccacctca ccctctcatc tcccattcac 1260
tcctacccac acacgaccca ctcccgtcca cctcacacca ctccaccaca caacatctcc 1320
tcccactcat cactacaccc tccacccact cacctccaat aactaccat catccacca 1380
accctccctt ccctcacatg cacacctccc ctactcccc cacctacaac cacctctctc 1440
acatccccct caaccaaccc cccaccatc accgctcga ctctctctcc cccacacacc 1500
ctccactaca tatccacaaa caaattaacc acaccagcgc accccacaac acacacacac 1560
ggtcacact cacacctcca cccccacac tccactctca ctctcacac tcacccctca 1620
ccaccccgca ccaccacctc ctctccccc tcccccca 1659

```

```

<210> 4
<211> 321
<212> DNA
<213> Homo sapien

```

```

<400> 4
tggtcgcggt cgaggtaact gatgcacaag gaaccgcatg ggactgggga tttacaaac 60
atacggacat gagagactca aacctcgatc ctggcacttc aaaatatgtt agtcacgtga 120
gcctgtggtg ggtgcccccc tctctcaatg gggggtgctg tttacagggt aacaattgat 180
cagaacaggt aaagggtaat taaaatgtat tcgccaaca aatgggcatt cttttataa 240
taatgaaaaa aacctctctg ttcgtaaaga gttgtctgca tcataaataa gaaaggatt 300
cataattata tcccaaaact g 321

```

```

<210> 5
<211> 1243
<212> DNA
<213> Homo sapien

```

```

<400> 5
tcctcaaata ctcacctcaa ggtcagcttc taggggagcc ccacccaaa tgccatgtcg 60
cctaggctct tctacacatg gagaccacaa ctgcatctgg acgtggaggt ttgcatggct 120
gtgatgctgt acgtgaagtg ctggcctgta gtacttggtg agtgctaact gttttcactg 180
tggttactca ttcattcacc cagaagttgg gcagctttct taaacaatgg gcaaagaccc 240
gcatgcagga gttagcaaga cacatgcagg caaaaaacgc acaagctagt ggggaagaga 300
gtatgagacc ctccccacca catacccagc acctggcccc tgccagaggg gcgatgcaaa 360

```

gtgggggtga ccacagggga gggggagaag caattgggtt ccaacctaca ggtttagtta 420
 agaaggcatt agttcaggat ttcacaacag attcactgtg cttggatcaa tcaacacatt 480
 gagagatgtc aggttcctag aggactcagc ctccacaccc ttcattctctc tctgcgggaa 540
 actttcaaca caattgatct aaatggcaca gtttttgttg atgccagttt gctagttcat 600
 aatcacgaac ccttagatga ctcatatgaa acaagagggtt ttttttcttc ttcttcttcg 660
 tagatgcaca aggaaccgca tgggcaagta tagattttct aaactatcct ttccctttgg 720
 ggccctggga aagcccaaaa atcagggaac aacagacact gactacctgc ataagctgat 780
 tgtaaaagcc agtcctgctt ctttcctgca ggactgggga ttttacatac atacgtacat 840
 gagagactca aacctcgatc ctggcacttc aaaatatgtt agtcacgtga gcctgtggtg 900
 ggtgcccccc tctctcaatg ggggtgctg tttacagggtt aacaattgat cagaacaggt 960
 aaagggtaat taaaatgtat tcgccaaca aatgggcatt ccttttataa taatgaaaaa 1020
 aacctctctg ttcgtaaaga gttgtctgca tcataaataa gaaaggatt cataattata 1080
 tcccaaaact gaaaagacaa caagtgtctc ccaagggtga ttgtgaaatc aactcgggtc 1140
 ttcaacgtac ttggaaagtc agccgggttt ggggttggtt ggggttggtt ttgggttag 1200
 agacggaggt tagctcttgt agcccaggct tcagtgcctg acc 1243

<210> 6
 <211> 392
 <212> DNA
 <213> Homo sapien

<400> 6
 tggtcgcggc cgagggtactg ctaatggggg attcttcttt gtctgatgta agggccaagg 60
 accaggagtg cagtctcttt gttttctttt atcttccttt agagtcaaatt gttaacagat 120
 cccaggttg ctctctttca ggagcctgga gaaggaaagt atcatgcagg gtcttctcat 180
 ccctgtctct tgctcaatta ctgtgacctt ttgtcccttt tttccccccc ataatttcta 240
 ttttcataat tttctctttg tctctatcct atttcttaag tctctctctt tttctattgg 300
 tctctttctt tctgtctcta actgtgtctc tttattgtct gtctgcctct gcattctctt 360
 tcctatttct gcctatcttt ttttttcttt ct 392

<210> 7
 <211> 254
 <212> DNA
 <213> Homo sapien

<400> 7

5

gttaaattga gaaacctgtt aacttacatt ttatgcggct ggccccttgg tatcacttac 60
 tgccagagat atttcctttc agcaccaggg cgaaagtctt ttaaattggct atggctctttt 120
 tcctactccg gtttggttta aaggcacatt aacggagggt tctcattagt caaagggtgg 180
 ctattcccaa aaaaagagaa aaacaaaaca gtggggggta ctcggcacaa cgttctcggg 240
 tgatgtccac atac 254

<210> 8
 <211> 1087
 <212> DNA
 <213> Homo sapien

<400> 8
 tggtcgcggc cgagggtactg gttcctcctt tttttttttt ttttttttgg aaatggagtc 60
 ttgctcctgt ccccgaaact ggattgcttg agtctcctt acctcactgg caaccttcca 120
 ccctgccgct gtcaaggcaa ttctccccc cctcaggccc tcccagaaga tagcatggag 180
 atttacaggt tgcattgcc aattgccaat ggcttaattt taggtattat tatagataaa 240
 aacgggtgggt ttcacaccca tgtgtttgag accacggcgt ggtctcctca aacctcctg 300
 tgaccctcag ggtgatctcc cactgtgccc tctcgagcct ctcccaaagg gtgcgtggag 360
 aataccagga catatgacca ccagtgggtg gtctgtgtg cctatatctc ctctatctc 420
 aatatctcta ttatatgtag ttatatattt ataactgtgt gtgtattata tactctctct 480
 actctctcat ctctctcact cactccactg tggtatctct ctattctctc aacaaatcac 540
 cacctatgtg tgccaagata tacatacaag atatagcccc cacgtgcgtt gcaagtgtat 600
 aacaacaaaa tatatacaca ctccctatta atattaaaga gtagacctag acatatatga 660
 tgctgacata tataacgcaa catagtgaac atcttcaata tacaggtgac gcacctataa 720
 acacacagaa agaggtgtga ctgtgctggt catcaccaat aatagagaaa tataattgag 780
 ggaacaaata ataatagaac ccagccgcct attttataaa cagaagatga tatcatctcc 840
 aacaaaacat aattatctac caatgtttac aactaggaaa ttacggccat tctaccacaa 900
 aagccgccga agataataat ataataaagt acgagaggac cacggtatga ggtagggtgc 960
 gtcaaccaat acagaacaat tatacacctc gttagagggt actgggtata ctgatgaag 1020
 aaagatatta tcatatataa ccacattcaa ccacagaatg ggtggcagta gagcaatatg 1080
 aggtagt 1087

<210> 9
 <211> 656
 <212> DNA

<213> Homo sapien

<400> 9

```

aaaaaaaaa gaggaattta cctaagggaa aaaataacta taaaaggacc aatttttata      60
ccaattttcta cagtttaggt aagggtttct ggtatattga ttaccttccc atttactatc    120
cctagctaac cgaggaaatgt ccaggtatc attctcccgg gatagttggg tacgttggag      180
tggaagggt tataaatttg gtttggccct gtggtagata atatgcatta ctaaagatgc      240
attttaagggt ccagggcgcg gggggcctca cgcctgtaat cccagcact ttggggagggt      300
ccgaggcggtg gcagatcacg aggtcgggag atttgagacc attccttggg ctaacacggg      360
tgaaaccctt gtcttacta aaaatacaaa aaaaaaatt tagccggggc gtggtgggcg      420
gggccccttg tttgtcccag gcttactcgc gggggctgag ggcgggagat tgggccgacc      480
ccggggggcg ggacgctttg ctcttgagcg gagattcgcg ccttggaattc caagcgtggg      540
cccgttggtg agtccgttc caaaaacaag gaaggtgctt taataccggg gtggcgggca      600
tacatttcgt ggtaacggt ggcggacctg gctaggtgcc tgggtgaatg tccgca          656

```

<210> 10

<211> 123

<212> DNA

<213> Homo sapien

<400> 10

```

gggatgatga tcatataggg cgaatggtea tctagatgca tgtcgagcgg ccgcagtttg      60
tgatggatcc aacacttcaa cactatttgt tttatttttc ttattaatat aagacggcag      120
gaa                                                                123

```

<210> 11

<211> 126

<212> DNA

<213> Homo sapien

<400> 11

```

acctctggaa aaggcaagga aacagattat cctgagagcc tccagaaaga atgccccttg      60
attttagccc aggagatcca tcttggactt ctgaccacaca gtaaaggtag aataaaaatg      120
tggtat                                                                126

```

<210> 12

<211> 274

<212> DNA

<213> Homo sapien

<400> 12

```

ttggaaaaaa tgtaaaattt cttatgtggg tgatttcaaa aatttgattt gaaatatata      60

```

ttaaaaaagtt gctatattgg cctattttaa attgctatca ttgatgggca gcatagtcaa 120
 tttcacaaag aaggccaaat tgtgcaaata ctaatatagt gggatgaccc tccttgggag 180
 agttacaaac ctcaatcaca aatgcaaaaa caaaaaatcc ataggcctac agagcagtaa 240
 ttttggttta ctagcaacca agaatatgat atga 274

<210> 13
 <211> 560
 <212> DNA
 <213> Homo sapien

<400> 13
 atagggagta caggcttgaa tatgtttggc ttagtttaga ttgtagatta ccaaggaaga 60
 atggcaattt gtaaaacaaa tttagctgct cagtatTTTT gagagaaaac tgaagagttt 120
 ttctcttgag gttttagaag cttttaagat tattagctcc ctaaacagat atgcatattg 180
 tcagtgatat cctaacattt tggaggttta atactattag gttaattata accaagaaat 240
 gtagaatgta gaatgaagca tttttatgc ctgaaatttg cttgtttgga aaaaatgtaa 300
 aatttcttat gtgggtgatt tcaaaaattt gatttgaaat atataattaa aaagttgcta 360
 tattggccta ttttaaattg ctatcattga tgggcagcat agtcaatttc acaaagaagg 420
 ccaaattgtg caaatactaa tatagtgggt gatccctcct tgggagagtt acaaacctca 480
 atcacaaatg caaaaacaaa aaatccatag gcctacagag cagtaatttt ggcttactag 540
 caaccaagaa tatgatatga 560

<210> 14
 <211> 356
 <212> DNA
 <213> Homo sapien

<400> 14
 cgagcggcgc ccgggcaggt actctggcct gggcaacaga ttgagactct gcctcaaaaa 60
 aaaaaaaaaa agaatgaggg gcaggaccca ggtgtgtgaa aagagggaca gataactgtg 120
 gtgtggtgtg gtggtggtgt aataagtctt attatcctat tggactttta aacctatgtg 180
 atttttttgc ttgtgaccaa gagggtaaatt tatttgacct tattaataat tcctagaaga 240
 aaacccttag aaaaaaaact cttctagact tgggacgagt caaagaattt atgattaaga 300
 cctcaaaagc aaatggcaac gaaaacaaaa atagacaaat tgagacttaa actaaa 356

<210> 15
 <211> 406
 <212> DNA

<213> Homo sapien

<400> 15

```

aagaacagaa agagagagag agagagagag agagagagac gagaccggga ggaaggcagg      60
tcgggaagga ccggcacagg gggccggacg gccggtaagg cgggggcaca gacagaaagc      120
aatgagtcga tagcgacaga ctgagagaaa gacaggaaga gagagcagag acagcacgac      180
aggtgggagg ccgggaggac ggacaaaaag aagacgacga ggaacgaaga acacgaacga      240
ccacgacaga aagacagaca cgagacgaaa ccgacagaca gaaaagaccg agagaaacga      300
caacacaaaa aaaaaaacia aaaaaaaagg ctgggggggtt atctggggac aaacgggtcc      360
cgggggaaat gtgtttccgg ccaaaccaaa atctctaaca caccga                      406

```

<210> 16

<211> 504

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (270)..(270)

<223> a, c, g or t

<400> 16

```

cgtggtcgcg gcgaggtaca agcttttttt tttttttttt tttttttttt ggcaaaaaaa      60
aataggcccg tttatttttt cctttggatc aaggggcact ttttgaaagc ctgtgggtgt      120
gccaagcttt ctcccaagg gggaggtatt atcgggggtt gggagcccaa gtctctcga      180
gggggggtgt aaagaggcac ctgggcaccc acacaagaga gcgcgaggag actctccaga      240
agcgccctac cctccatata tgtggggcgn ggaacaactc acacgcgcgt tggggcgat      300
aaacctcggt gtgggctcat ataagcagct gtggtgttct ctgctgtgg tgttggtgta      360
gcaacacaat ggttgggttg tgtactatcc tcgcgggctc tcacacaaat gtttctccac      420
cacacacaaa cacaattaag gcggggaaag caacacagat ttagaaaacc acccgagtc      480
cacccaagaa ctcaccaaca aatc                      504

```

<210> 17

<211> 234

<212> DNA

<213> Homo sapien

<400> 17

```

atgactttct gaccatatag gccatggtca ctaatcatgc cgagcggcgg catatgtgat      60
ggattggtcg cggcgaggta ctacactctc ttggttacca tagttttata caatagtaag      120

```

ttttaaaata gagaaatgtg agttatcata cttcattctt tttttaaaga ttatttggt 180

atcctggggtt ccttgcaatt ctatgtgaat tttagaattc gccagttaat ttca 234

<210> 18

<211> 16

<212> DNA

<213> Homo sapien

<400> 18

taaataaata aataaa 16

<210> 19

<211> 132

<212> DNA

<213> Homo sapien

<400> 19

agctgggcaa tgtggcaaaa ccctgtctct actaaaatac aaattttgct gggctctgtgg 60

gctgcccttg tatcccagct actcaggggc tgggaggaga ttgcttgact tggaaggcgg 120

gtgcctgtgg ta 132

<210> 20

<211> 445

<212> DNA

<213> Homo sapien

<400> 20

gagatgaacg actcactatg gcgaatgtgc ctctagatgc atgccgagcg gcgcagtgtg 60

atggatggcc gcccgggcag gtactgggat tacaggcatg agtcacgggtg cctggccttc 120

tccagatat ttaaaagtag gggtcacgga agctagttag tctctattag ttcttgaact 180

gataaaactg atgaggaaaa aaaaaaagaa atagaccac tcagagacaa agagataaga 240

atccagtgtt ggcccaagcc agagagagag agagagagag agagagagag agacgacaga 300

atgaacgccc gaacgccctg gtggagggtc tctgaattt agggcacact aagatgttcc 360

tagtcctaaa tgatccccct ttctccctcc cctagactg gttctaagtg gatctccttt 420

tgcttgacc aatagagtga aagtg 445

<210> 21

<211> 681

<212> DNA

<213> Homo sapien

<400> 21

tggaggcaga gtctcactct atcaccagc ctggagtga gtggcacagt ctcagctcac 60

tgcaacctcc acctcctggg ttcaagcgat tctcctgcct cagtctcctg agtagctggg 120

actacaggtg tccgccacca tgcctggcta atttttatat ttttagtaga gacagtgttt 180
 tgccatgtgg gccaggtgg tctcaaactc ctgacctcag gtgatccacc cacttcggcc 240
 tcctaaagta ctgggattac aggcattgagt cactgtgcct ggccttctcc cagatattta 300
 aaagtaggggt tcacggaagc tagttgatct ctattagttc ttgaactgat aaaactgatg 360
 agggaaaaaa aaaagaaata gacccactca gagacaaaga gataagaatc cagtgttggc 420
 ccaagccaga gagagagaga gagagagaga gagagagaga cgacagaatg aacgcccga 480
 cgccctgggtg gaggttctcc tgaatttagg gcacactaag atgttcctag tcctaaatga 540
 tcccccttcc tccctcccc tagactgggt ctaagtggat ctccctttgc ttgcaccaat 600
 agagtgaag tgaagctttg tgttcaacc aacccttctc agttgccaag cactgtgcta 660
 gttctggatg aacagcagta a 681

<210> 22
 <211> 516
 <212> DNA
 <213> Homo sapien

<400> 22
 caggctacaa tagcaaacac acagaactat ttcctgctct tgcctaatg gggttcaaaa 60
 tgacttgctt tagtgctatt aagagttata cattcagaca aaaatgtgca tgagtgctaa 120
 cttgggatat ccaggtgctg ctacaggtgc tagatatcga acagtccaca agaattctgc 180
 agttcctgcc ctcaagaagc ctacctgcc acctgttaat ctacctggca ctgttctggg 240
 gtgtgaaggt atggagacaa ccaaggotta gcattctgcac atggagtta caaacaactg 300
 gaagtcata atttcacccc cataagtaag aaatagtaat aagtgtttta ttgggtatct 360
 attatgtaaa atgccttata catagcaggc attttcctaa gtccttttta ggtattatct 420
 tacttaagtt tgtaacctac cccatggcat aggccaccaa cattagccca gttttgtaaa 480
 ggaggaaacc tgcgacagag aggaatcaac tgactt 516

<210> 23
 <211> 514
 <212> DNA
 <213> Homo sapien

<400> 23
 gaagtcaggt tcaaggcgct ggcgtccag ctgatccctg gacctgaaca gggacctgtt 60
 ccctgtcctg cttggaagtc ccattcctggg tgtgggcagc cagagaaagg aagcgttctc 120
 ccagtgtgc catgggctgc agccctaccc tgctgggctg agtccggtgt ttaaggagg 180

11

```

gaggagggag gagaggggtg aagagctggg ccttctggta gctttttata attatttcta 240
aaatgctata tttggatatt attttctgct tctacaaata aaacatgcat atgtgtaaaa 300
aaaattcaac acatttaaag aacaaaaaca acaaacaaaa agaaaaaaaa agggcgctgt 360
gggggtgtac ccctgtgggc caaaagcgcg tgtgtccccc gtggtgtgtg agcaattttg 420
tgttctctcc gcgccctcca atattcccc ccaaaattat tagggaaaaa cacaagggcg 480
ggtggaccct cgctcaccat aactgatag ctgc 514

```

```

<210> 24
<211> 668
<212> DNA
<213> Homo sapien

```

```

<400> 24
ccgcccgagc aggtctgttc tcagcagcag taagagcctg gtcaatctga accttctagg 60
caatgaattg gatactgatg gtgtcaagat gctatgtaag gctttgaaaa agtcgacatg 120
caggctgcag aaactcgggc ttcaaaaaga cctgcacaat gtagtgagag aggagataca 180
gacctcacag aaggagctct gtctgaaact caagtgtgcg tgggatttta atgaccttga 240
agacaagtgg tgggtggtgat ccacagatt agatgccacg tggcttgacc atggatcttg 300
ggggaaagcc accaggacat cctggcctgt gtgtcgctcc aatgtcacca tttgtgggga 360
caaatgagct gttccctgca ggaggctttg tcacggttgt tggaggccgc ccattgcacg 420
cccaggctcg gaatcctagt gtaatactgt gtctggtacc aagatcataa gttggtgtg 480
ccttcagtct tgtctatgtc ctcttggtg taatgttttt aattcttgga ggtgttgaga 540
gaattcaata aagcaaagca tataaaagta aaaaaaaaaa aaggaaaaaa aaaaaagcc 600
gtgggggtaa ccaggggggc agaggcggtc ccggggggaa agtggtttcc cgcccaaatt 660
tccacaat 668

```

```

<210> 25
<211> 755
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (190)..(190)
<223>

```

```

<220>
<221> misc_feature
<222> (190)..(190)
<223> a, c, g or t

```

```

<400> 25
gcccgggcag gtgaaattat aggattttca accattcccc tagaatgggc tggctgcctg      60
acttagggac cacgttttgc cagaatgtta agtttgaagg tcatggcccc attgttctga      120
agtatctatt ggagaaacag aacatcacgc ttagtcggct ctgggacaca aagataaatt      180
atatgacgcn cacacatgcc tagctgagaa gtatgatagg agcttcaggc gctggctcta      240
ggttcgtgga gttggttggtg ggcctatcgt cgattgttaa tctcatcttc taggccgtcg      300
agacgtcatc aacaattaat ctcttggtgg gactcagtgt aaagcctctt aatcacgctc      360
gtttttacgt tcatagacat cttttttcct ccgtctgaaa taatgagata gagattcttg      420
tctcctctgt aggacttttc ttctccccgt cactcccaag acttgagtta ggtgcattcc      480
tagtatcgag atactctatt gtaattttctg ttttcctgta gatatttcca tagtcataga      540
cctgtttgcc tgtagataga aattctgcct tattcgtgat tcgacgcttc agctotttgc      600
atagcgtcta gcccatggta gacactcagt aatcactgac tgagttaaag aatagaatag      660
acctaaataa tataaaagca aaaaagctgg gggtagcagg gccgagcggc cccgggggga      720
atgggttaccg gggccgaatc cccgaaagaa aaacg                                     755

```

```

<210> 26
<211> 1137
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (190)..(190)
<223> a, c, g or t

```

```

<400> 26
gcccgggcag gtgaaattat aggattttca accattcccc tagaatgggc tggctgcctg      60
acttagggac cacgttttgc cagaatgtta agtttgaagg tcatggcccc attgttctga      120
agtatctatt ggagaaacag aacatcacgc ttagtcggct ctgggacaca aagataaatt      180
atatgacgcn cacacatgcc tagctgagaa gtatgatagg agcttcaggc gctggctcta      240
ggttcgtgga gttggttggtg ggcctatcgt cgattgttaa tctcatcttc taggccgtcg      300
agacgtcatc aacaattaat ctcttggtgg gactcagtgt aaagcctctt aatcacgctc      360
gtttttacgt tcatagacat cttttttcct ccgtctgaaa taatgagata gagattcttg      420
tctcctctgt aggacttttc ttctccccgt cactcccaag acttgagtta ggtgcattcc      480
tagtatcgag atactctatt gtaattttctg ttttcctgta gatatttcca tagtcataga      540

```

cctgtttggc tgtagataga aattotgcct tattcgtgat tcgacgttc agctctttgc 600
 atagcgtcta gcccatggta gacactcagt aatcactgac tgagttaaaa aagaaagaaa 660
 gaatgaaata aatgacttaa tgggtattaat catacaaaca acagctttaa acaacagtga 720
 acctcttgaa catccaaatt ttttttttta cttcttgagt gcaatactga cactagagaa 780
 gcctaaaagg taagagaata taccctctt atttcacaca ggatggtagt aacaataata 840
 gctaaaatga tggggtgctt agtgcagaca ccatgaatga agacactctt atttaatatg 900
 cacaaaatcc ttgatacaag tataattaac atcatcattt tatggacaaa aacctgagt 960
 ttttagagttt ctaatctggt tcaacatttc acagctaggt gagcaatgaa gtctgggtgg 1020
 tgaccaatct gacacccaaa ccatacgtaa tgggtctcaa gcccactc tacagccaac 1080
 tcagggtctg aactacgact ccagttctaa acgttgctcc atacctctag cacattt 1137

<210> 27
 <211> 15
 <212> DNA
 <213> Homo sapien

<400> 27
 taaataataa ataaa

15

<210> 28
 <211> 123
 <212> DNA
 <213> Homo sapien

<400> 28
 aggggaatgag ggacggaaag agagagacag cgaagagagg agaaagagtt tcagaatttg 60
 gcaaaggctct caaggctcaa gctgggtgtc tgaagccttt caaaccccca gttttatcac 120
 cac 123

<210> 29
 <211> 3426
 <212> DNA
 <213> Homo sapien

<400> 29
 atgggtctatg gggcttccga ggcgatcagg ctgtgtcagt cttcagccgc taagccaaga 60
 aggagtcagt cagagagcct tgggccagag ttccaggggc tctgggagtg gctgccaggg 120
 agaagtgaaa tgacaacctc actagataca gttgagacct ttggtaccac atcctactat 180
 gatgacgtgg gcctgctctg tgaaaaagct gataccagag cactgatggc ccagtttgtg 240
 cccccgtgt actccctggt gttcactgtg ggccctcttg gcaatgtggt ggtggtgatg 300

atcctcataa aatacaggag gctccgaatt atgaccaaca tctacctgct caacctggcc	360
atttcggacc tgctcttcct cgtcaccctt ccattctgga tccactatgt cagggggcat	420
aactggggtt ttggccatgg catgtgtaag ctccctctcag ggttttatca cacaggcttg	480
tacagcgaga tctttttcat aatcctgctg acaatcgaca ggtacctggc cattgtccat	540
gctgtgtttg cccttcgagc cgggactgtc acttttggtg tcatcaccag catcgtcacc	600
tggggcctgg cagtgctagc agctcttcct gaatttatct tctatgagac tgaagagttg	660
tttgaagaga ctctttgcag tgctctttac ccagaggata cagtatatag ctggaggcat	720
ttccacactc tgagaatgac catcttctgt ctcgttctcc ctctgctcgt tatggccatc	780
tgctacacag gaatcatcaa aacgctgctg aggtgcccc aataaaaaaa gtacaaggcc	840
atccggctca tttttgtcat catggcgggtg tttttcatct tctggacacc ctacaatgtg	900
gctatccttc tctcttctca tcaatccatc ttatttgga atgactgtga gcgagcaag	960
catctggacc tgggtcatgt ggtgacagag gtgatcgct actcccatg ctgcatgaac	1020
ccggtgatct acgcctttgt tggagagagg ttccggaagt acctgcgcca cttcttcac	1080
aggcacttgc tcatgcacct gggcagatac atcccatcc ttctctaaa acaactacc	1140
atcagagaat actacaaaca cctctacgca aataaactag aaaatctaga agaatggat	1200
aaattcctgg acacatacac cctcccaaga ctaaaccagg aagaagttga atctctgaat	1260
agaccaataa caggctctga aattgtggca ataataca tcttaccaac caaaaagagt	1320
ccaggaccag atggattcac agccgaattc taccagaggt acaaggagga actggtacca	1380
ttccttctga aactattcca atcaatagaa aaagaggga tctccctaa ctcatcttat	1440
gaggccagca tcatcctgat accaaagctg ggcagagaca caaccacaaa agagaatttt	1500
agaccaatat ccttgatgga cattgatgca aaaatcctca ataaaatact ggcaaaccga	1560
atccagcagc acatcaaaaa gcttatctac catgatcaag tgggcttcat ccttgggatg	1620
caaggctggt tcgatataca caactcaata aatgtaatcc agcatataaa cagaacccaa	1680
gacaaaaacc acatgattat ctcaatagat gcagaaaagg cctttgacaa aattcaacaa	1740
cgcttcatgc taaaaactct caataaatta gaattggaaa aaactacttt aaagttcata	1800
tggaacccaa aaagagcccg catcgccaag tcaatcctaa gccaaaagaa caaagctgga	1860
ggcatcacgc tacctgactt caaacttaca ctataagact acagtaacca aaacagcgtg	1920
gtactggtac caaaacagag atatagatca atagaacaga acagagccct cagaaataac	1980
gccgcatatc tacaactatc tgatctttga caaacctgag aaaaacaagc aatggggaaa	2040

```

ggattcccta ttttaataaat ggtgctggga aaactggcta accatacgta gaaagctgca 2100
actggatccc ttccttacac cttatacaaa aattaattca agatggatta aagacttaaa 2160
cgtagacct aaaaccataa aaaccctaga agaaaaccta ggcattacca ttcaggacat 2220
aggcatgggc aaggacttca tgtctaaaac accaaaagca atggcaacaa aagccaaaat 2280
tgacaaatgg gatcgaatta aactaaagag cttctgcaca gcaaaagaaa ctaccatcag 2340
agtgaacagg caacctacaa aatgggagaa aattttcgca acctactcat ctgacaaagg 2400
gctaatatcc agaatctaca atgaactcaa acaaatttac aagaaaaaaa caaacaaccc 2460
catcaaaaag tgggcaaagg acatgaacag acacttctca aaagaagaca tttatgcagc 2520
caaaaaacac atgaaaaaat gctcaccatc actggccatc agagaaatgc aaatcaaac 2580
cacaatgaga taccatctca caccagttag aatggcaatc attaaaaagt caggaaacaa 2640
cagggcaccc ctccgacccc cgctgcgagg ggtttccgcg gggaggtgga cgaagcgtgg 2700
gccaggacag gcgcctggca ctggggtttc agagccgcga gtaggcctg aaggccggga 2760
gggaccgcca gtccccactg gaagccgctt ccgtgcagcc cgggacagag tttcagaatt 2820
tggcaaaggt ctcaaggctc aagctggttg tctgaagcct ttcaaacccc cagttttatc 2880
accacatatg agctcttagc atatgaaact gcagctagga ctaaaaccag aggaaggcac 2940
acagaatctt gactgctctt cagaaataca tggcactcca ttaaaatgaa gatttttgct 3000
gctcactcca gaaatattga atcaaacgt ggagcaggag acaggcaagg ccagtgaagg 3060
agtgtggctc cgataggatt gacaactact accttagatg atgaatcaag ctcagaaggg 3120
atccctctgg atttccctgc ggtgaagaaa ccgaggcaat gaacaaagta atatacccaa 3180
gatcacattg ccagtgaatg ggagagctga cgtttcaatc tacacagaac ctgtgctctt 3240
agctatctaa ccgctttact tgggaagtgat gtgagattaa aaaaagaaga aaaacaaaat 3300
attttcttat gctttcaaaa agttcaaaaat taatcaaggg aaccgtttct ccatggggac 3360
aggagcttct ggaaggctgg acccaatcat tacaggctca gtccagggcc tttccttcac 3420
accaac 3426

```

<210> 30

<211> 259

<212> DNA

<213> Homo sapien

<400> 30

cccaagtagc tgggattaca ggcacgcacc accatgcccg ggtaaggggc tggctctcag 60

caccaggacc tggcacagca tctgacctcag aggacagctc agttcatgat tgccagatgg 120

16

ctgcacgcag cgcgaggag agccccctgg gtctgaaata gaggtggag aggaggactg 180
 tgagtccatg aggagggaga gtgatggat cccacacaca agccagcgtg tctaggactc 240
 ctatctgaag aactgcag 259

<210> 31
 <211> 948
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (284)..(566)
 <223> a, c, g or t

<400> 31
 aactcacata taggcacgc tctctaata tgctcgagcg ccgcaagggt atgatggatg 60
 tcgcgccga ggtacgagaa gtcttcttag actgtttaga gatactctgt gttctataac 120
 atgcgagcat gagatccatg ggctaataagg aaacaattgc atggtatata ataacaaaat 180
 gtttcattgt gtatatcttt tacagaaggg totgagtatt cacctgagtc attatcctgg 240
 ttttctgagt agtgaaattt acaaatcata aaaattgaat gctnnnnnnnn nnnnnnnnnn 300
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
 nnnnnnnnnn nnnnnnnnnn nnnnnnaaaa aattgaatgc ctgtgctcag aggtagggtta 600
 tgggtggactc accctggggt gggcagccac gcaatgcaat ggaatgagaa caaccacagg 660
 cccaccccca agagcagccc aggtcacacc tgggcaccca ggggtctctg tggagccttc 720
 tgaggctgac ccaagataaa ttcacattaa ccaggactgt tacagaaaaa aaacttcagc 780
 agaattaaag gagtctaact gagcatggaa cgatgggtgt atggggagcc ccaaaattaa 840
 gcagattcag atgactcccg ggggtgctcat ggtcggaaca tttttggttt aaaacaaaaa 900
 aaaacagggtg ggttttcggg ctgtgttctg gtgattttcg caacaaaa 948

<210> 32
 <211> 545
 <212> DNA
 <213> Homo sapien

<400> 32
 atttgatgt tttagagaag ggagtcttgc tgtgttgccc aggtttgagt gcagtggcta 60

ttcacaggca caatccttagc acattgcagc ctctggcctc aaatgatcct cctgtctcag 120
 gcccctgagt agaggggact atagtgtgtg tgccactgca tccagctgca ctcttttaac 180
 ataccagttg gtttacatat tttcagtgga ctcataaatg tcatagtttt tttttagtag 240
 ttttaatttg tttaaatcag gatccaaata aggttcacaa attggaattg attaatatgt 300
 cttttaagtc tctgaagttt tcccttcac tttttttctc cctcgtaact tatttgtgga 360
 agggatcagg ttgtttgtct tatagcgctc cccactgcc aagattttgaa gggttatattc 420
 tccacagtat agttttacgc cttcttctgt ggtttctatt cctataatgg gtaatggtct 480
 agaggtagtc ggtggtgctc cggttgggtt ttgggcactg cttgtgctgt gtggggccct 540
 cggtc 545

<210> 33
 <211> 912
 <212> DNA
 <213> Homo sapien

<400> 33
 ttatatcaca cttttcattt tattctgcct tccaggaaaa tttcttgggt gggcacctta 60
 ctttacagtt tacagaaggc tcacacaaaa tctcatgtta gcctcacac gctttgatac 120
 ccacttcaca gatgtgagac ccgaggctcc cagctgagtt cacaacaccc acaagtagca 180
 gaggagtatc tccaggccgc ggcttcctgg ccaacacatt ttcttcaact cctgcttttg 240
 ggttcccttc cagaccacag tctgactcac tcttcctga catgaaaccc gaagtgcct 300
 ttctccaaag ttagaacaag tagttactgt atctctccga caagcaaca ttctttcagc 360
 aaacattggt tttgtttttt agagaaggga gtcttgctgt gttgccagg ctgaaatgca 420
 gtggctattc acaggcacag tcatggcaca ctacagcctc tggcctcaaa tgatcctcct 480
 gtctcaggcc cctgagtaga ggggactata gtgtgtgtgc cactgcatcc agctgcactc 540
 ttttaacata ccagttggtt tacatatattt cagtggactc ataaatgtca tagttttttt 600
 ttagtagttt taatttggtt aaatcaggat ccaaataagg ttcacaaatt ggaattgatt 660
 aatatgtctt ttaagtctct gaagttttcc cttcatcttt tttctcctc cgtaacttat 720
 ttgtggaagg gatcaggttg tttgtcttat agcggtcccc actgccaaaga ttttgaagg 780
 tatattctcc acagtatagt tttacgcctt cttctgtggt ttctattcct ataatgggta 840
 atgggtctaga ggtagtcggt ggtgctccgg ttgggttttg ggcaactgct gtgctgtgtg 900
 gggccctcgg tc 912

<210> 34
<211> 380
<212> DNA
<213> Homo sapien

<400> 34
tatttcatgt taatgggctg ggagattagt atgttaagat ggttgaatat tgtcccttca 60
aatcatggt gaaattactt cccaaaattg acagtattga cgagggtggg cctctcacgc 120
aaggatgatg gtcgatgagg tctgctgtca taactggata acaaattggg gtctgctgtg 180
gaatggttat gggtatgctg gatgggattg tgggtcctaa gaagaggaga gagacctgag 240
taaattctcag aactgcatt tttgctgga ttatttcgat ctgagaaact tacagagaaa 300
cctcccatgt gcctactgtg cttgactttt taaaataacg ttgtttaaaa acacaaaaac 360
atgggtcgtc gtctgatgtc 380

<210> 35
<211> 714
<212> DNA
<213> Homo sapien

<400> 35
gccgggcagg tactaattta aacactgaca gccacctaaa tgacagcaat gggcacttgg 60
ctgaggtggc cacacagcac ggaatctcag gggctccaga gccagaaaca gcagggtcag 120
tttcaaattc cagctcctct attcactgtc tgtgtgacct tgagcaaac attgcctcca 180
aacctgtttt cttacctata gaaagtggat gataataaca gtattttacct catagaaatg 240
tggctaggaa taaattagat gatgtgtgta aaatgtagcc cataggaagt gttttataag 300
tgttttctat tattaataaa attatgctat aatcatgat aaattcattg caaacataaa 360
aaggatgctg tagaaaacta atgacaatgg aaaatgttta tggatattgg ttatgtgcaa 420
ggagggttata ttactgcagg atatccattt ttgctaaaaa atatattttt cagaggggaa 480
aaacagggtga aataattttt acatcaaat gctaaccata gttattttctg actgatggag 540
ttatagggga ttcttcgttt tattatttg ggtttttcta tatttgcaa aaaaaaaaaa 600
aaaaaaggct gggggtaatc atggccatag ctgttcctg tgtgactttg tttcccgccc 660
aatcacatcc ccaaaaaaaaa aagtccaac aaaccaacaa ccaccacaaa aaaa 714

<210> 36
<211> 474
<212> DNA
<213> Homo sapien

<400> 36
ccaggaaaaa gaaaaaaaca aaaaaaaaaa aagcgcgtgg gcggtactcg gggccatagc 60

tgtcccggt gtgaattggt tactccgctc cacaattccc ccacaccatt gcgcaacaa 120
 cggccgaccg aaacacaagc aaagaaacag gagacgcaca caaacagggg acacggagag 180
 aaacaaacaa cacaccgacg gcaacaacca aagccaagga acacacagag caacaaagag 240
 gaccagaaaa agaacgagcc acgcgaaagc agagaaaaac gaacgaaccg gaacaagagc 300
 agagacggag acgcccgaga gcgagcgag acacaaaacg ccacagaaac gcacgaaagc 360
 ggaaacgcgc acacgacgaa gcgagacgac aaaacacaga gcaaaaaagc gaaaaagcaa 420
 agagaacaag acaaccata cagaagagaa acagacaaaa agagaacgac aaca 474

<210> 37
 <211> 914
 <212> DNA
 <213> Homo sapien

<400> 37
 acacacatgt caaaataatc aagttgcata ccttgaatgt aaacaatddd tatttttcaa 60
 ttatcttcaa taaagctggg gaaagaaaat ctcataggt ggttttgttt tgtctttttt 120
 tttttttaag atgcagtctt gcctccttgt tcaccaggt cgtggattgc aatggcggtt 180
 gactctcagg cttcactgtg caaccctcgt ggctccgtg ggtttcatag caattctctc 240
 cgttgctca gccctcgtgg tagtacgtg ggatttacag ggagcccgcc attcaccgcc 300
 cgagctaatt tttgggtatt ttctagatag agagtgggt tttcgccatt gtgtggccag 360
 gcgtggtctc gacactcgt gactctcgag atgatccacc tgtcttcagc ctcccaaaag 420
 tgctgcgatt tacggggcgg gagccaccag gcctgggcca ccttgagggt attattaatg 480
 gcaatctggc ccggggccag tctttagca gcttttgggt acccattttt ttggccccc 540
 ttctcatttg gtgcctatt tggatagata tttggactca tttccatttg ggctatagta 600
 catccttaag tctcctccta ctacctgat acggtccaat gtgatcttgt ggctttcccc 660
 tttgaataca tcaagcatgc tttcactctc accatctttt gcaattgtgt ccttcagtca 720
 aaaatactcc acattaaggc atgattactg cattgcttgg ggcagtagtc agcacctggt 780
 ggctaagaat caccttaaca aggcctgcct ctaacatcca gaagttatac agtttctctt 840
 ttctgcccgg gttagatggt taggggagtc tatgaacgac ctgggtcttg acacctcaaa 900
 aggttactgg gaga 914

<210> 38
 <211> 923
 <212> DNA
 <213> Homo sapien

<400> 38
 gcgtgggtcgc ggccggaggt acaagaggaa agagcaggag atggaatgag atttaaataa 60
 atgagaagac tgacactgat ctgaatagca atatgaactt cagtgggaga aagggtcttcc 120
 tcctagacaa tggggaaagt cgtaagtagt ttgctgtttt aaagttctat cccagattc 180
 tgtgatgtcg tggtatatatt ccataatcat ttagtgaaga caggcactaa ctttgaagg 240
 tgggtgcaca ggagatgagg aacaatacat acggggacat atcaccagca ctatcctgct 300
 ctatattggc ccttcctgtt tctatggagt tccttgcat ctcaacagta gtgcagtct 360
 tgatcttcaa gctgtgacct cctgaccag caaacaagag agtgtggctg gttcacagg 420
 cagtgagacg agtgcgtggc tccgaagagc tggcaaggaa acaatggggg agaagacca 480
 ctgcccccca tccaacacac acacacacat acaacactag tagtagctac actatacaca 540
 cacacacagg agtttgtgtg agagacaagc tctctttatt tctcccaat ataatagata 600
 tagattatatt ctttaagagag aggaagaaa agtggcagt gtatcactta atcactcagt 660
 gctatatgtg acaatattgt tgttgaacgc tatatttaaa tagattcaga actacattgg 720
 agaatgtcct cagataggac gagtctaag gaattccaat ggaagagagg gaggttcaat 780
 gtgggcacaa actcccagag ggaccccggt agaataagg agattatatt tatgcgcct 840
 taatacacta gaagtcagtt taaaatgcc cggtttggaa ggacaaaggc cttgggcctg 900
 gggaatgttc ccaaaaatgg tcg 923

<210> 39
 <211> 576
 <212> DNA
 <213> Homo sapien

<400> 39
 cggcgcccg gcaggtacac caggcctggc taatttttaa attttgtaa gatgggtcta 60
 gctgtgttgc acatgctggt ctcaaactcc tggcctcaag acatccctgc cctgcctctc 120
 gccttgccct cctaaagtga tgggattaca ggcgtgggcc accatgccca gcctgaatct 180
 tttttttttt ttttttttgg gagatataga tattgtaaat attttaaca agcatctgta 240
 ggtaacaga tttgaacgcc tctcctaggc cactacaaat tgaccctca ggcaggggct 300
 ggctgccaca gggctgcccg tgcccccat aggtaccag gggttgaggg caaatctgcg 360
 gcaggggggc tctgggggga gcaggtgggt gacccattt gaccagctt ttccatttaa 420
 aggggattaa caccctgaaa aacacaggaa accacaacaa aaacaacaaa aacaacaaa 480
 cggcgggtgg gggataatca ctggggcaca taagctgttt cccgggggtg aaaatttgtt 540

21

ttccccccca aattcccaca aaatataaga aaaagc 576

<210> 40
 <211> 734
 <212> DNA
 <213> Homo sapien

<400> 40
 cccacagaga gctgtagggg atttttcttt gtttaactag agagcacagt gtttggcata 60
 tggcagcact cacactggta ttcttccttt agagcttcct acatttgctc tggtaataag 120
 cagcagaggc aggagtattc tagagccttg gggcacagga agctgggtgt ctgacagggg 180
 tcacctctga ctatccaagg tgatgctgga ggagtgtggt gcttcttggtg ttgcttggtc 240
 tcttttgcac cttataacat gggaacagca aagagctata aagggcattg gaggcctgga 300
 tgctggtggc tcatgcctgt aatccacaac caacatccct tgggggaagg ggccccgaac 360
 gggcggtgtg cgttgagaat tccaaacccc ttgatgggtc aggagacgtt taatgaacta 420
 tgtcttggcc aacatggttg aaaaccccc caggtttcct tttcctttaa accttttaaa 480
 aaggaaccag aatttaatct aaaaagaaga aaaaattttt ctttgggacc cttctggcgg 540
 cgggtagggg ttggaggatt tggggccctg gttagtatgc cccttggtta aaatacccca 600
 tgctactcca tgtgatgact tgaatggcac gtgatgacat ttgcttttga aaacctgggg 660
 acgggttggg cacagggttt tgccatggtt ggaaaatggg ccccccccg ggggaaaaaa 720
 cacacgcggg ctta 734

<210> 41
 <211> 604
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (511)..(511)
 <223> a, c, g or t

<400> 41
 cgccccggga ggtggaaatt cagagtagtg aaattgttag cagttgttct taaaatctgg 60
 gctctcattt catggcgatg tctttggtct ttgtgttgaa agcaggacct ctgtaagccc 120
 cttctcctcc ccagactgtg agtttggtgg gtgttatgac aacggtgagg ggacggggga 180
 gggcccctcc aggaagtgtg catctcagtc cagtgcgggg tcagcgagta aaggacttac 240
 taggttggcg acctgagtgt caccagagc cagagaagtt tccatatctc aatgaacctt 300
 ttggattcga agagagatca ttactaactc cacggactgg ccttagaaga ctcttctct 360

gacatcatcc aattcattct gccacataaa gataggaata aaaagaaaga acaaaagaag 420
 ggctgggtgcg gtctaacgag ggggtcatagg cgagtcaccc gtgggtggaa attttgttgt 480
 gccgcgccaa ctctatcgcc ctccaacatg ngagaggaca agaagggggg cttttgcct 540
 tttgttacct ttgaccgaca ttgactcccg tggataagat tgtcctccag aattaccata 600
 ctgg 604

<210> 42
 <211> 898
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (493)..(493)
 <223> a, c, g or t

<400> 42
 agacagacaa aaagaaaaac aaagcatgac gcatatgggg actgggcatc taatgtgtgt 60
 cgagcgggcg agtgtgatgg attggctcgcg gcgaggtaca atcacaggct ccctgaagcc 120
 tcaattttctg agctcacgtg agcctccttg cctcagcctc tatcatcaga gcagggtgt 180
 cggttatgga ctcagagtgc tgagattata ggcgtgagcc aagccgtgcc caatcgtctt 240
 aatgcttttg ttctcagttg tggtcgtcac gcttgatatgt aatcgagaag ttctcacacc 300
 tgtcaactct tagttggcac accatagttt tctctagagg tctactgtat ccttggtatt 360
 gttggttcaa gagtgtataa tatgtcaaat cagctgtcgt ttctgtagat cgccgccatc 420
 ctctcaaagg tgttagaaat tatccgcttt cgtgtatggt tagaaataat taaaacttta 480
 agacgggtatc ttntgcata gaacgttttg taggattgag aaacatttaa aagaattatc 540
 ctctcgtatt aacaaatcaa tcggtttcag aataaacata aagaaacaag ataatagaga 600
 aaagcgctct ggggggtgaa ccgcaggggg ccacctggag cgtgtgtctg cccgggggg 660
 gggacattgg gttatcgccg gttcagcatt tccgggtccac ctattagtgg ggagacccaa 720
 aaaagttccg gtgggataaa gattgtcatt ccagaaaata acccattacc tgtgaaatgg 780
 gcaccaactg tgaaaagttt aagaaaagcc cctgttcgaa aggcacgacg atgggctagt 840
 ggcttcatcc atgccaaga ggtcggaagt tggttctggg acacttttgg gtgggtgc 898

<210> 43
 <211> 408
 <212> DNA
 <213> Homo sapien

<400> 43
 cacaacatac gagcatacga gcatggggag aaacacgctt tcacaaatga cgccaagatg 60
 agaagaggac acgcacacga acatctaacc taccattatg aacagagtaa ttagcagcac 120
 agtcaagatg actgacaaag cagtagatca acagacagta ataccaagaa cgcaaagagt 180
 taatgtatcc tagatagatg gaacaagtca atgggaaatt agacgaactg atgagagtaa 240
 aaacagtaga agtaagaaat agtaaaagaa gaactaagtc aatagcagac aagaaacaga 300
 acgaatagaa aggacagagc acaagccaag catagaagca agaagcagca catgcaagac 360
 aagaaggaca gaagacagat aaaaatcaag atagatacat acagaaca 408

<210> 44
 <211> 804
 <212> DNA
 <213> Homo sapien

<400> 44
 ggccgcccgg gcaggttgta atcccagcta cttgggaggc tgaggcagag aattgcttga 60
 acccgggagg cagaggttgc agtgagtcca gatcgtacca ctgcactcca gccaggcaac 120
 agaaggagac tccatctcaa aaaaaagaaa aaaaggtaag gccggactca gtggctcaca 180
 cttgtaatct cagcacttcg ggaggaggct gaggcaggca gattgcttgc gcttaggagt 240
 tcaggactga actaggcaac atggagaaac catgtctcta caaaatataa aaaaattagc 300
 tggacatggt gtcttgacc tgtagtcca gctactcagg aggctgagct gggagtatca 360
 cttgagccca ggaagtgcag attgcagtag ccaagatcat gccactgcac tccagcctgg 420
 gaaacatagt gagatcctgt ctcaaaaata ataataataa aataggccga gcgcggtggc 480
 tcacgcctgt aatcccagca ctttgggagg ccaaggcggg tggatcacga ggtcggagat 540
 caagaccatc ctggctaaca cggtgaaacc ccatctctac taaaaatata aaaaattagc 600
 ccggtgtggt ggtgggcgcc ttgtagtccc agctactagg gaggcggagg caggagaatg 660
 gcgtgaaccc cgagggtgg agcttgagc gagccgagat tgcaccactg cactccagcc 720
 tgggtaatac agcgagactc catcccaaaa aaaaaaaaaa aaaagctggg ggaaccgggc 780
 aaacttcccg gggaatgttc gtca 804

<210> 45
 <211> 1146
 <212> DNA
 <213> Homo sapien

<400> 45
 gcggccgccc gggcaggtag taaaaatata aaaattagcc aggcacatg gcggacacct 60


```

gtaatcccag atgattgggc agctgaggca ggagaattga ttgaactcgg gaggcggagg 120
ttgcagtgag atgagattgc gccattgtgc tccagcctgg gcaacaagag cgaaacttca 180
tcacaaaaaa aaaaaaaaaa aggaatttac taaggaaaaa ttaattatta aaagacattt 240
ttattccatt ctgaggttag taagggtgtt cggttttttg atttactttc ccattttacc 300
tattccctag ctaaccgggg aaatgtggtc ccagttatct cattccttct ccgggggtag 360
tgtgggatag tttgggatgg aaatgtgctt attaaatttg ggttgtgggc ccgtggtgga 420
tagattaatt atgtgcatta cataagaagt ggcattttta tgaggcccgg ggcgcgcggg 480
tggccttccg cgcccgtggt aatctccccg agcactttgg gggagaggcc ccgggggcgg 540
ggcggaaactc cgcgagggtt ctggggggag aattgggaga accatctccg tgtggctata 600
accacggggg tgaaaaccct gtgtgtccct atacttaaaa aattaccaa aaaaaacaaa 660
tttaggccgg gggcgtgtg gggggcgggg gccctgtgt tttgctccc ggagttaacc 720
taaagagaga ggcatagaag ggcaggggag agaattgggc acagaagccc ccgggggggg 780
gcggggacgc ttttgcaagt agaagcgggg agaattcggg cgccatttgg acttccaagg 840
cctggggggc acccagggag tgggagacc cccgggttcc cccaaaaaac aaaaaacaag 900
gcgaaagaat gcgcgtttaa atttaccctg gagggacagg ggggaccggg ccaattaaca 960
aattttcctc cagagggggg ttaaaaagg cagtgggggg gaaaaccagg gggccaaaaa 1020
aggggtgtcc cccggggggc tgaggatggg tctcccgccc acaaacaca aggagaaggc 1080
aagagaacgt aacagcagcg cgccaatagg agaaaccacc agacacggat aacaagtaac 1140
caaacg 1146

```

```

<210> 46
<211> 160
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (16)..(16)
<223> a, c, g or t

```

```

<400> 46
cacaacatac gagcantacg agcacggccg aaggacagag acgaaacgag agcaaaagga 60
acaaagaaca gaatacaca gaaaggaaag ataagaaaga gagaagaaga aaaagataag 120
aaaaaacaaa agaaaaggaa aaagagcaga acacgagaga 160

```

<210> 47
 <211> 993
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (221)..(221)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (255)..(255)
 <223> a, c, g or t

<400> 47
 ccggccccggg ccggttaattg ccaccagcga aagcgtacta ttgatgcctt ccggccagga 60
 gcccggtctt tcctgatctc atcgctgctc ttctcgcttg cgtcgctctg cttgaagacg 120
 actgcagggc ccttgcagac agtcttcgat aaactttctc catcttctaa catcccgagg 180
 ccatagtcgg ctgtcttggg aagtccgaac ttgcttagtg nctagagtgg agagcatcat 240
 cgcgtcgctc tacanctcag agacctcta gctaaagtgt ccatcaacct cttcttttagt 300
 tgagcatgga gagaacatgc ttgagagaaa gtccaagag gtatgaggta tgacctttga 360
 gaagatacac tgtgatgagg tttgactaga ttagtggata gcctatctat taaatcatgg 420
 cctggaggta acaatgtgca aaactgaagg agagagccat atccataaga gtagttaaca 480
 ctatgatccc cttgcctgtt gcgctctgac caaatatagg aactttaata ctaccttgta 540
 acctaagaat agaaatcaac ggatggccat tagtgggcaa actgggataa acactataaa 600
 agaagaaaaa caaccatatg tgaaaagata aataacaagg agaaaactag tgttaaaata 660
 aaagaccgga gaaagtagct gaaagcgcaa aatacgggag aagtgacaaa aagccgcgaa 720
 aaaaagagcg cccttaaggg tgaaaggctt cccaggaat tttcaacaca aaaaaaaga 780
 aaagagaaga gaaaaaaaca aggggggaaa aaaaaaatgg gggaaaaccc cattttttat 840
 aacacatat gtggaatagg aagaaaaaag aaaaattaaa aaaggaaagt acagggggaa 900
 aaaaacaaaa gcccccaaca aatatttcgt aaaaaaaaaa aaagagggac atgtggtgga 960
 acacaaaaaa aaaaaaaaaa aaaaaaaaaa aaa 993

<210> 48
 <211> 498
 <212> DNA
 <213> Homo sapien

<400> 48

26

tggcgttaat ccaaattgggt ccataggctg gttttcccct ggtggttgaa aatttgttta	60
ctccccggct ccaccaatt tcccaacaac aaaccatacc ggaagccaag cggagaaaac	120
caccgaagca cccccgcga caccgggcac caccacgcgg gcaagaccat cgggccagg	180
acagggcggc gcacgcgcag caccgagagc gagcgaaagg gcgagacca gagcgacgca	240
gacgagagcg gacggcgaaa agagcggcga cagaaaagag aaaaagcaga gagacacgaa	300
gacacaaaag ggaagggcag gggcgtgaga ccccgagggc gggccccgag agaacaagag	360
acacgcaggg gcggggggta gcgccacca cgacaaaatg ggaggggcga gagagagacg	420
gcagagacaa gagcaaaaga gaggggagac agcaggggcaa acagaacaag tggcgtggca	480
aggacggagg cagcgcag	498

<210> 49

<211> 905

<212> DNA

<213> Homo sapien

<400> 49

gcggccgagg tactggttcc tccttttttt ttttttttgg aaatggagtc ttgctctgtc	60
gccgaagctg gagtgcattg atctctactc actgaacctc caccgcggcgg ttcaagcaat	120
tctcccacct cagcctcccg agtagctggg ataacaggctg cttgccacca tgccgggcta	180
attttggtat attagtagag aggtgggttc caccttgttt ggcccaggcg ggtctcaaac	240
tccgtgacct cggggtgact ccacctttgc ctgggcctcc caaagttgcg tgggattacg	300
ggcttgacct aggggtgccg ggctatctc cttcatctct ttagtgtaac ttgaacgggg	360
tttaatctct ctctctctca cgtgtttcct tctcacattt cacttggtga gttactcaaa	420
tagcccaggg tgatgttaca gatttacgtc cttataacaa gggaggcata tatggcttta	480
cacatgcttc tgaggtggcc ttacaaaagt gtcgggtcca taggataatt gactatgcac	540
cttttaaaat atttcaatat ccattacagt tagctccac ccagtattat aagacttatg	600
taccaagcgt tatcttgggg tcatggatat ctacctatca tgtgctgttg gtttatgacc	660
atataattcg tgtgtacccc ttatttcccg gtgaacactc tgttgaatt ggtgacttgg	720
gtctaagaaa cagtgttaat tttggaaagt attccggttg accttgacaa ctaccctgct	780
ttcataatat tcctgtccct atttaattat tggccctttt taaaattcac gtagcttttt	840
taaaacttct cctttacttt gatttcaccc cagggggggg tccccttttg cccggtgtgt	900
taacc	905

<210> 50

<211> 698
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (289)..(367)
<223> a, c, g or t

<400> 50
gcggccgcgcg ggcaggtgcc agcgcagggc tttctgctga gggggcaggc ggagctgagg 60
aaaaccgcgt atgagttttg tgtctctttg aaagatagag tattaactca acaactactt 120
acaaaaaata tagtccagag gttactaaga tatgctgagc gttacgttag cacacgtaat 180
tcaatagctg aagatttgac gagaatcata ctgcaaagac ttacaagagt agcctgagga 240
aggagaagat actgggtttg ctaggacaca tgacggaggc tgagatgann nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnngga gcaggtcacc tgggtcggag ttagactagt ctggccaaca tggtgaaacc 420
ccgtctctac taaaaataca aaaatttgct ggggtgtggtg gcgtgtgcct gtaatcccag 480
ctactcagga gactgaggca ggagaattgc ttgaacctgg gaggtggagg ttgcagtgag 540
ccgaagtgtt gccattgcac tccaacctgg atgacaagag caaaactttg tctccaaaaa 600
aaaacataaa gaaaaaaaaa aaaaaaaaaa aaaggtgggg gggaaaccat gggcacaaac 660
gggtccccgg ggggaaattt gtttcccgcc caaattca 698

<210> 51
<211> 406
<212> DNA
<213> Homo sapien

<400> 51
gcccgggcag gtacaagctt tttttttttt tttttttttt tttttttttt tttttttttt 60
ttgggaaaaa aaaaaagggc tctttttttt tcttggccca gggggctttt tgagaaaccc 120
gggggggtggc cacttttccc caaaggggtt gattttgggg tggggccccc ttttcgcggg 180
gagaaggagg cgcggtcgcc acacacgacg aggagatcac aagacgtctc cccacaatat 240
gtggggagga gttacacgct ggtgggaaac aacgtgggac aacgactgtg tccgtggggg 300
agaaatgggt cttcccgccg acaaactcca ccacaaaatc atcagaagaa aaggggtatc 360
tacacaaaac gacacagaac ctccgcgcac atcacgagaa gatatg 406

<210> 52
<211> 725

<212> DNA

<213> Homo sapien

<400> 52

```

gtggtcgcg cgaggtctcg tgagccccct agaccatcac ggatgccgag cttcgggtaa      60
ctctcacagt ggaaggttcc cagccgccc ctaatcccg tcgaagcagc cctgagaaac      120
atcgcccatt ctctctccat atccccccc aaaaattttg ccacccaac acttcaacac      180
tatttgtttt atttttctta ttaatataag acggcaggaa tgtcaggcct ctgagcccaa      240
gccaagccat cgcaccccct gtgacttgca cgtatatgcc cagatggcct gaagtaactg      300
aagaatcaca aaagaagtga atatgctctg cccacacctta actgatgacc ttccaccaca      360
aaagaagtgt aaatggccgg tccttgcttt aagtgatgac attaccttgt gaaagtcttt      420
ttcctggctc atcctggctc aaaaatcacc cccactgagc accttgcaac cccactcctg      480
cctgccagag aacaaacctt ctttgactgt aattttcctt tacctacca aatcctataa      540
aacggcccac ccttatctcc cttegtgac tctctttttt ggactcagcc cgcctgcacc      600
caggtgaaat aaacagccac gttgctcaca aaaaaaaaaa aaaaaaaaaa aagcctgggg      660
gaaccggggc aaagcggccc ggggggaaat tgtttccgcc caatcaaaga aaaaaaaaaag      720
gggag                                             725

```

<210> 53

<211> 968

<212> DNA

<213> Homo sapien

<400> 53

```

ggtcatactc ctattcacgg ttctcaacta ctcatcatg ccctgctctt gtttacctg      60
ccggtttaca ctgtttttcc aagccatcac agctgatatc tcctgggtgct atccccaaac      120
tgccaactctt aactcttgaa gtaaataaat catctttgct ggcaggacta tgctgaatct      180
ccttaggcac tctctaatac gacatcctga gtcgtcccaa ttcttagacc ttttatacct      240
gtttttctcc ttctgttatt ccatttagtt tttcaattca tacaaaaccg tatccaggcc      300
atcaccaatc attctatatg acaaatgttt cttctaacat cccacaatc tcacccctta      360
ccacaagacc tccttcagc ttaatctctc ccactctagg ttccacgcc gccctaatac      420
ccgcttgaag cagccctgag aaacatcgcc cattctctct ccataccacc ccccaaaaat      480
gttcgcccgc ccaacacttc aacactatct tgttttattt ttcttattaa tataagaagg      540
caggaatgtc aggcctctga gcccaagcca agccatcgca tcccctgtga cttgcacgta      600
tacgcccaga tggcctgaag taactgaaga atcacaaaag aagtgaatag ccctgcccac      660

```

29

```

ccttaactga tgacattcca ccacaaaaga actgtaaatg gccggtgctt gccttaactg 720
atgacattac cttgtgaaag tccttttcct ggctcatcct ggctcaaaaa gcacccccac 780
tgagcacctt gcaacccccca ctctgcctg ccagagaaca aacccccctt gactgtaatt 840
ttcctttacc tacccaaatc ctataaaacg gccccaccct tatcttcctt tcgctgactc 900
tctttttgga ctcagcccg ctcacccag gtgaaataaa cagccatggt gctcacaaaa 960
aaaaaaaaa 968

```

```

<210> 54
<211> 679
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (393)..(393)
<223> a, c, g or t

```

```

<400> 54
cgccggccgg cgccggtgct cctgcaggaa taccgctaaa agcaccggag aggactgaga 60
tgtggatggt gcgttttggt cacttacggt ggcgtctgag ctccatggtc cccccaagta 120
gtgagctgca gccccgcatg gagaagagct ctgcgacggt tcaaccatac ggtaatgcca 180
gtgcgcactc agaccttgcg agcgtccccg cgaaccgtct cgtacacagg attcgtcctg 240
ggtcacctgg atacctgtag ttccttaaga catcgactgt gattgcgcca tgtggataga 300
acaggtaatt ttgctcactg cggtgtcgca tatatagttg ccaaactatg acgctgtcg 360
tttggtagac gctggtcaca tgctgtata ggnttggcct gttaaagtctc tggtcggtac 420
cgtggtgtgc ctggtagtc tttgtctgtg tatcgcgaca tctggttccg acgagcaagc 480
agtatggtag tcaaagaaac tccgggaaac gaaatgaact gcaaggcaga ttggattacc 540
cggatcctgg aagggtgat gcagaataag gatgaatggt agagggattg gaaaatgtct 600
ggttcaacta acgcctctac ttggtaatca cgctgagggt agaatagggt ctaccctccc 660
cgaaaccac aaaacaaag 679

```

```

<210> 55
<211> 1618
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (408)..(408)
<223> a, c, g or t

```

<400> 55
 gccgcgcccc gatgctgaac cgcccgactg agccgacagt gccggcgctct tggctcgattc 60
 ggtccgatga tgctctcggg tgcgtctaca gcattggetg gtctactgg tagacagagc 120
 taggcacgtc atgcgtacat cccgattcac tcccagatcg tgtgacagaa atgcacgctg 180
 agtagagtgc ttcgtagaca gacgagaatt tctaataac taaagtgtcg tgaccaatga 240
 cctttctggc cagctaacgg tacgcactat tgtgacaacg cttggagacg gcacatagtt 300
 ctagccottg actaagacgc tgggacgata gatcagtgcg gtcatacgca tgtcaccgtg 360
 tgttgactct tcgtggatcc gctcgtacga attcctacgc gcacaacnat tacgggcaag 420
 ccaaaagcgc aatcgcgctc gctcacgcat agcttgtccc agtgatgctg tatgatggcg 480
 taatatccct gcatcacatt gcactaacag tgacgtgcat tcggatccac gggaaaatcg 540
 aggacaacct acaagggtgca agcacagcgc agttatgggtg caggagaggt aaccatatta 600
 acggcgattc caataaccga ataattccgc ggaggccacc gacgttgata tacacgtgcc 660
 attgtcacag atctaaaagt gaggacagcc atataggggtt agggcacgct ttcggagttg 720
 cacottatct cgtggatat atcaccatga tgtgggtaga cactttaaag tttgcgccct 780
 taacacgcca cttactttta ttccgtccct taaggaggac ctttataaag aacccatata 840
 tataggaggg cggatacaca taacaagaga ccggttcaca caacaggaca gaataaagta 900
 caacgaactt ccccggtttc gggaacgcag gctaataaaa caggcgtag tggcggttaa 960
 catccaatga ggctccaatt aagggttggt tatccccctg ggtgcttgaa agcaagcccg 1020
 gtattcattc cccggggtcc caccaaaatg tccaccgaca acagacgaag aggaacgaca 1080
 aacagacaac gagaaccgaa cgagtgacca aagcagagca acacagaagc caaggaggac 1140
 cacgaaaaca aagataagaa gacagacgca gcaacacgaa agacggagat aagaaccaa 1200
 gaggaaacac aaagggaat caaccacgaa acgacaaaca agcaaaacag aacaccaagc 1260
 ccaacaaaca aacacaaaaa gcaagagaga aaaacaccaa gaaaacaaa gaaacaaaac 1320
 aaaaacaaac gacgcaggcg aacacagaaa cgagcgacaa aatcagaacc agacagcaca 1380
 caggcgagga cgaaccgacc acgccaccaa ctccgaaccc gcgagcacca gcgagaacac 1440
 cgcaaaagca agaagcgaag ccacaacaca gaccacgagg acagaccaga gagagcaaag 1500
 agcgaaagca gacgagcgcg aacacgcgag cgcagaagag aacagaagaa gaaaaagaac 1560
 acacagaaac aagacgacag gcaagaagca ggaaagacga gccgcccacc cacgccgg 1618

<210> 56

<211> 1875
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (359)..(359)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (459)..(459)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (515)..(515)
 <223> a, c, g or t

<400> 56
 cgtgcgcccg ggcgtgtgac caggacttcc tggttttcca tacttaactg tggccacagc 60
 gatgccggta tcggtctatg agtagacatg tgagtctagg cccctgtga gtccaagaa 120
 tctgctctat acagcatcaa gtcgtggccg tcgactagcg tagtatctgg cgactcaatc 180
 cctcatagac caatggttca cactgatatg tgaatgctag cgaattacgg tgctgcgggc 240
 tacgtgggtc atgggcagaa gtgacctggg cgcaccgggc atgagcgatc gggtagatcg 300
 actaagacta tggctcgcgc gtgtgcccac cgtggaccca agacagaagc acgtaacang 360
 cacaatgggt cctattccgg tgtcatcgag ggcttgtgga tccagcgcat gatatagtag 420
 gccaaagggtg gcgatctcgt tgggacgcga tgtgagggnc cttcgacgtc gcaattccat 480
 gtgcagacgt ataacgtctt gtagttctcc aatancgcat agatatatac acatggatag 540
 ttggaacaaa ttgcttatac atcacggcct catgcggggg ttgtcacaac caagcgtagg 600
 caaatcaggg gaaaccggca aatccccgcg ggggggggtg ttagagcacc gtgtgggtgt 660
 gtatatcctc cggagggcgc tacacgacag agttttctcc cacacacaca gaaattttcg 720
 gcgacagaag caacacacca caggcggagc agaaagccag gagcagcaca aaggagcagg 780
 agagaaagaa acacaaagaa gtagaacaga acaacgaaca acaaagacaa aataacataa 840
 caagaaaagt aagtaaaaaa gagagagcag cacatagaag caggtcacac gacaactctc 900
 agagagcaca ccgtacacac agtacaaacc cacaagaagt acaaagaagc aaaagacaac 960
 acagagaggg taaatcagaa gagcaacaca acaagtagaa gggaaagcac aaacatcaca 1020
 aagcacaggc atagataaaa cgcgaaacaca gaaaactgca cagtaagaga gcgtagtact 1080

32

aaacgacgag	ccacagccga	ccagcagaca	ggcgacagct	agttacaaga	gggacgggag	1140
agcacacgcg	aacagacacg	agccgagcaa	tgcacacccc	accagcagca	cagcacacca	1200
aagagtcaca	aaaaaccgtc	gacgacaagt	cacacacaca	cacacaggta	gcacaggcaa	1260
aagacacaga	cgagcagcag	gcaacactaa	agacgagagg	aagaatatac	acagacaggc	1320
aagacacaac	aacaaacgaa	caaacgaacg	ggacataacc	aacagagaaa	gtagcgaaaa	1380
accacacaga	agagaaacac	agaaagaagg	agcacaatac	gaaggaccgc	gagagagaga	1440
cacagcaaaa	gaagaagaga	tagtcagcaa	gccagagcaa	aacgacacac	acagaagcaa	1500
caacaagaac	gccagcaagc	cacagagcac	agcaggaaaa	aatagcagaa	cacagacaag	1560
aagacagcga	ccaaaagaag	gacgccgcaa	aaggaaagac	gagaagcacc	cacacacagg	1620
gcaggacaaa	acacacaaag	agaaagaagc	cagaactcac	acaagcgcac	gagaggaagc	1680
gcgaagaacg	acaacaaaca	taaagataag	agacgacaag	ggaggagaag	tgataacacg	1740
cgacaaggag	caaggcgaac	agtgaagga	gcagcggaca	agaaaaggga	caagagaagg	1800
caagacggac	agaacgacag	agaggaaagg	caacagcagc	aagcagtaga	agggcagtcg	1860
acaacacaaa	gcact					1875

<210> 57

<211> 781

<212> DNA

<213> Homo sapien

<400> 57

gcgtgggtcgc	ggcgaggtct	cgtgagcccc	ctagaccatc	acggatgccg	agcttcgggt	60
aactctcaca	gtggaagggt	cccacgccgc	ccctaatacc	gctcgaagca	gccctgagaa	120
acatcgccca	ttctctctcc	atatcaccac	ccaaaaatct	ttgccacccc	aacacttcaa	180
cactatcttg	ttttattttt	cttattaata	taagacggca	ggaatgtcag	gcctctgagc	240
ccaagccaag	ccatcgcatc	ccctgtgact	tgacgtata	tgcccagatg	gcctgaagta	300
actgaagaat	cacagaagaa	gtgaatatgc	tctgccccac	cttaactgat	gaccttcac	360
cacaaaagaa	gtgtaaagtg	cgggtccttg	ctttaagtga	tgacattacc	ttgtgaaagt	420
ccttttcctg	gctcatcctg	gctcaaaaat	cacccccact	gagcaccttg	caacccccac	480
tcctgcctgc	cagagaacaa	accctctttg	actgtaattt	tcctttacct	acccaaatcc	540
tataaaacgg	cccaccctta	tctcccttcg	ctgactctct	tttcggactc	agcccgcctg	600
cacccagggtg	aaataaacag	ccacgttgct	cacaaaaaaa	aaaaaaaaaa	acaaaaaaag	660
gcgtggggaa	ccctgggcca	aagcctgtcc	cgggtgttga	aattgtttct	ccgtcccaat	720

cccattatattt gacacaaaca atcgtaaaaa aacgaaacaa aaacacaaaa ccataaaaaa 780

a 781

<210> 58

<211> 5434

<212> DNA

<213> Homo sapien

<400> 58

atggctgaga atacaaattc aggtaagatg aagagaaaaag cccacctgca cccacgtgaa 60
 ataaacagct ttattgctca cacaaagcct gtttggtggt ctctttacat ggacacacat 120
 gaaatttggt gccgtgactc ggatcggggg acctcccttg ggaaatcaat cccctgtcct 180
 cctgtttcttt gctccatgag aaagatccac ctacgacctc aggtcctcac accgaccagc 240
 ccaaggaaca tctcaccaat tttaaactcag cggcaagtcc cgctttcctg gggcaggggc 300
 aagtaccctt caaccctctc tccttcaccc ttagcggcaa gtcccgtttt tctgggggag 360
 gggcaaacca tcacggaccc cgagcttcgg gtaactctca cagtggaggt ggctgccgct 420
 gcattaatac ttttagaggg cctcaaaatc acaaactatg ctcaactcac tctctacagc 480
 tctcataatt tccaaaattt attttcttc tcacaccgga cacatacact ttctgtctcc 540
 cggctccttc agctatactc actctttggt gagtctccca caattaccat tgttcctgggt 600
 ctggacttca atccggcctc ccacattatt ccggatacca cacctgacct tcatgactgc 660
 atctctctga tccacctgac gttcacocca tttcccata tttccttctt ccctgtttct 720
 caccctgac acacttggtt tattgatggc agttccacca ggctaataca cactcacca 780
 gcaaaggcag gctatgctgt agtatcttc acatctatca ttgaggctac cactctgccc 840
 ctctccacta cctctcagca agccgaacta gttgccttaa ctcaagccct cactcttgca 900
 aaaggactac gcgtcaatat ctatactgat tctaaatgtg cctttcatat tctgcacat 960
 catgcagtca tatgggctga aagaggtttc ctcactacac aagggtcctc catcattaat 1020
 gcctcttcaa taaaaactct gctcaaggcc gctttacttc caaaagaagc tggggctcatt 1080
 cactgcaagg ggcataaaaa gccgtcagat ccattgctc taggcaatgc ttatgctgat 1140
 aagggtggcta gacaagcagc tagctotcca acttctgtcc ctacaggcca gtttttctcc 1200
 ttcacttcgg tcaactccac ctactcccc tctgaaactt cccctatca atctcttccc 1260
 acacaaggca aatgggttctt agaccaagga aaatatctcc ttccagcctc acagggtccat 1320
 tctattctgt cgtcatttca aaacctcttc cacgtaggtt acaagccgct agcccgtctc 1380
 ttagaacctc tcatttcctt tccatcatgg aaatctatcc tcaaggagat caottctcag 1440

tggtccatct gctattctac taccctcaa ggattgttca ggctcctcc cattcctaca	1500
catcaagctc ggagatttgc cctgcccag gactggcaga ttgactttac tcacaagcct	1560
cgagtcagaa aactaaaata tctcttagtc tgggtagaca ctttactgg atgggtagag	1620
gccttccccca cagggctctga gaaggccacc gcggtcgttt cttcccttct gtcagacata	1680
attcctcggt ttggccttcc cacctctata cagtctgata acggaccggc ctttattagt	1740
caaatcacc c aagcagtttc tcaggctctt ggtattcagt ggaatcttca tatcccttac	1800
catcctcaat cttcaagaaa ggtagaacgg actaatggc ttttaaaggc acacctcacc	1860
aagctcagcc tccaacttaa aaaggattgg acagtacttt tacctcttgc tcttctcagc	1920
atcacagcct gtcctcgaga tgctacaggg tacagtccct ttgaactttt atatggacgc	1980
actttcttgc ttggcccaa cctcatcca gacaccagcc ctctaggcga ctatcttcca	2040
gtcctccagc aggctagaca ggaaattcgc caggctgcta atcttctctt gcctactcca	2100
gatccccagc catacgaaga caccctagct ggacaatcag ttcttgtaa gaatctgacc	2160
cctcaaactc tacaacctca atggaccgga cctacttag tcatctatag taccctgact	2220
gccgtccgcc tgcaggatcc tcccactgg gttcaccatt ccagaataaa gctgtgtcca	2280
tcggacagcc agcctaatcc ctctcttcc tcttgaagt tgcaagtact ctccctact	2340
tcccttaaac tactcacag gtccagcaag acttccccag acatttcaca tcagcaagct	2400
gccgcctcc ttcacactta tttaaaaaac ctttctcctt gtattaaact tactcccccc	2460
atatttggac ctctcacaac acaaactact attcctgtgg ctgctccttt atgtatctct	2520
cggcagagac ccactggaat tcccctgggt aatctttcac cttctcgatg ttcctttact	2580
cttcatctcc gaagcccaac tacacacatc actgaaacaa ttggagcctt ccagctccat	2640
attacagaca agccctctat caatactgac aaacttaaaa acattagcag taattattgc	2700
ttaggaagac acttgccctc tatttctctc catccttggc taccttcccc ttgctcatca	2760
gactctctc ccaggccctc ttctcgttta cttataccca gccccaaaa taacagtga	2820
aggttgctcg tagatactca acgttttctc atacaccatg aaaatcgaac ctccccctct	2880
acgcagttac cccatcagtc cccattacaa cctctgacag ctgcctccct agctggatcc	2940
ctaggaatct gggtaacaaga cacccttctc agcactctc atctttttac tttacatctc	3000
cagttttgcc tcacacaagg tctcttcttc ctctgtggat cctctaccta catgtgtcta	3060
cctgctaatt ggacaggcac atgcacacta gtcttcctta cccccaaaat tcaatttgca	3120
aatgggaccg aagagctccc tgttccctc atgacaccga caccgacaaa aagagttatt	3180
ccactaatc ccttgatggc cggtttagga ctttctgcct ccactattgc tctcggtact	3240

ggaatagcag gcatttcaac ctctgtcacg accttccgta gcctgtctaa tgactttctct	3300
gctagcatca cagacatatc acaaacttta tcagtcctcc aggcccaagt tgactcttta	3360
gctgcagttg tectccaaaa ccgccgaggc cttgacttac tcaactgctga aaaaggagga	3420
ctctgcatat tcttaaatga ggagtggtgt ttttacctaa atcaatctgg cctgggtgat	3480
gacaacataa aaaaactcaa ggatagagcc caaaaacttg ccaaccaagc aagtaattat	3540
gctgaacccc cttgggcaact ctctaactcg atgtcctggg tcctcccaat tcttagtctt	3600
ttaataccca tttttcttct tcttttattc tgaccttgta tcttctgttt agtttctcaa	3660
ttcatccaaa accgtatcca ggccatcacc aatcattcta tatgacaaat gtttcttcta	3720
acaaccccac aatatcacc tttaccacaa gatctccctt cagcttaatc tctcccaatc	3780
taggttccca cgcctgccc aatcccgctt gaagcagccc tgagaaacat cgcccattct	3840
ctctctccat accaccccc aaaaattttc gccgcccac cacttcaaca ctattttggt	3900
ttatttttct tattaataga agaaggcaag aatgtcaggc ctctgagccc aagccaagcc	3960
atcgcatccc ctgtgacttg cacgtatacg ccagatggc ctgaagtaac tgaagaatca	4020
caaaagaagt gaaaaatgcc tgccccacct taactgatga cattccacca caaaagaagt	4080
gtaaatggcc ggtccttgcc ttaactgatg acattacctt gtgaaagtcc tttgctggc	4140
tcactctggc tcaaaaagca cccccactga gcaccttgca accccactcc tgcccgccag	4200
agaacaaacc ccctttgact gtaattttcc tttacctacc caaatcctat aaaacggccc	4260
caccttatac tcccttcaact gactctcttt tcggactcag cctgcctgca ccaggtgaa	4320
ataaacagcc atgttgctca cacaaagcct gtttggtggt ctcttcacac ggactcacat	4380
gaaatttggt gccgtgactc ggatcggggg acctcccttg ggggatcaat ccctgtcct	4440
cctgtctttt gtcctgtgaa aaagatccac ttacgatctc agggcctcag acccaccagc	4500
ccaaggaaca tctcaccaat tttaaatcag gacccactg aaaatcggac tgttcaactg	4560
cacctggcag ccaactgcag agccctgaa accctggccc aagggtctct gactgactcc	4620
ttcccagatc ttctcagctt agcagctgaa gactgacact gcccaattgc ctcggaagcc	4680
ccctagacca tcacggatgc cgagcttogg gtaactctca cagtgaagg ttcccacgcc	4740
gcccctaact ccgctcgaag cagccctgag aaacatcgcc cattctctct ccatatcacc	4800
ccccaaaaat ttttgccacc ccaacacttc aacactatct tgttttattt ttcttattaa	4860
tataagacgg caggaatgtc aggcctctga gcccaagcca agccatcgca tcccctgtga	4920
cttgcaagta tatgccaga tggcctgaag taactgaaga atcacaaaag aagtgaatat	4980

36

```

gctctgcccc accttaactg atgaccttcc accacaaaag aagtgtaaat ggccggtcct 5040
tgctttaagt gatgacatta ccttggtgaaa gtccttttcc tggtcatcc tgggtcaaaa 5100
atcacccccca ctgagcacct tgcaaccccc actcctgcct gccagagaac aaaccctctt 5160
tgactgtaat tttcctttac ctacccaaat cctataaaac ggccccaccc ttatctccct 5220
tcgctgactc tcttttcgga ctcagcccg ctcgaccag gtgaaataaa cagccacgtt 5280
gctcacaaaa aaaaaaaaaa aaaacaaaaa aaggcgtggg gaaccctggg ccaaagcctg 5340
tccccgggtg tgaaattgtt tctccgtccc aatcccatta tttgacacaa acaatcgtaa 5400
aaaaacgaaa caaaaacaca aaaccataaa aaaa 5434

```

```

<210> 59
<211> 1106
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (364)..(364)
<223> a, c, g or t

```

```

<400> 59
gcggccgccc ggcaggtagc cacatactac ccgccagagg ctgacccgcg ctagatagag 60
taccctacta cccagatagc tagtaccatg ggctatctac tagatagaag gcggctggct 120
caccaggcta cctgcagacc aaaaacgaca tgcagcgaat catggctgac tcttatccaa 180
aataacttaa cttacgtcct cagccaagaa tgccatatgg accgcacgctc tggcaagctt 240
acccttctct acattgggac ggaaataaaa atgatggcat atttatatgc aagtggcgag 300
acattggacc agagagggaa ggggaacctc agcacgggag accgcagcaa tcaaccgaac 360
cacnccttta ccgggtttgt ggggttatgg gggattgggg tgtttggtgt gtgtgtgatt 420
gtgttttttg ggttattttc atttttgggg ggggttgcca acaaaccaac agcgaataaa 480
cacgagacag tagcagacgc ataagacaac accgaccaa ctagaggcag acagcaccga 540
cgaaaccggc aaagccgaca tcaacaataa aaaaacacaa ctacaacagg cacgacccag 600
acaaacaaca aagatcaaaa cacacaggac aacaacaaag aaacaaaagc aacaatagaa 660
cgcacaaacc aacacaaaga aaaacacata aaataagacc atagatagcc aaggaaagcc 720
cagagcacga cagacccgag gataacgcga gggagacagt ttcagctgag ctgagcggag 780
gaagtggccg agggggagcc gccgctcgcc aggtgggtgt ctggagtata ggtatgacga 840
ctgcagggct gatggagcta agtgctgcgc ggttcggaaa cgcgtaacgc ggagtaagcc 900

```

37

gcgatgagtg gcggaggctc gggtcagagg ctgaccacgc gcgtaccgta ggacccagcg 960
 ggcaagggcc gaaggaagat gtttgcaaga acccgtttgg ggcgacaccc aagaaagcag 1020
 tcggagtaga gaaggaacca ggaagagcgc gagaggtgac ttatcatgaa tgactcaggt 1080
 ttggcgaaga cgagggaaag cacgca 1106

<210> 60
 <211> 122
 <212> DNA
 <213> Homo sapien

<400> 60
 cacaacatac gagcagagag cgaggccaga agacttggcc tgagtccgga agttagtcgt 60
 ctgatctgac gttctacgta cttgtatattt tattgaagga ctgatgagcc ctgctacctc 120
 cc 122

<210> 61
 <211> 929
 <212> DNA
 <213> Homo sapien

<400> 61
 tggctctgttc tgtgctgctg tgctgatgct ggtatcatgc tcacgtcaaa tgtgctggta 60
 atactgtggt atcacacaga acatcatgtg ggataactgc agtgaacaga acagtgtgac 120
 gcaacaagtg ccatgcaaga atggccatgt gacacctgta cacaacggc ctgcgtatga 180
 ctggccgtaa taatccagca acaagctata ctgacgcaca acaccgcaa acacacaaaa 240
 agcacacaaa cggaagaacg acacgagcac acaagcaagc agcacaacaa gcacgcagcc 300
 aacagaccag acaagcaca gagcaaccca ccaaagagac agacaacaca acagagagaa 360
 gaagacaacg caacgcagaa cagaaaaacg cacaagaaag ccagcagaag cagaaacacc 420
 cgaaaggaac agaaagaaaa gcagaaagga acgaaacaaa agaagagaga agacaagaag 480
 agaaacagca cacaacgca gacaaggaga gagaaagaaa gactcaaacg agcagagaaa 540
 caaagacggg agacagagga gaagagacaa ggacagcgaa gagacagaag aaagaaacaa 600
 agaagcagac aacggccaga gaggacgaga agacaaacag gcgaagaaga caggaagaga 660
 cgaaaacgac gaagaagagg acagcagaga acaacgcaga gagaagaaaa aagaagaaga 720
 gagacggaca gcaggagaca gaaaggagga acaaagacaa acgagaggag cagaacaaga 780
 gagacaagct acgaccgacg agcgaagaga gacacaagca agaacaacag agagcgacgg 840
 gcacacgcag agcagcgagc agccaaggag acaagagaag agaacagaga cacgacgaga 900
 aagaataacg acacgagaag aagagagga 929

<210> 62
<211> 598
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (270)..(270)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (330)..(330)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (332)..(332)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (334)..(334)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (533)..(533)
<223> a, c, g or t

<400> 62
ggcgacggga aaaatgactc atcatatagg gcgactgggt ccactagatg cctgctcgag 60
cggcgcgagt gtgatggatt tttttttttt ttttttttcc tttttttttt tttttgggtt 120
tttttttttt ttaaaaaaaa aaaaaaggcc cgtaattttt tcctttggcc cactggggcc 180
tctttttgaa accctgttgt tgttcgcaa cctcttcccc caaaggaggg gattcctccg 240
ggtgtgtggg acgcacaagt ctctctcagn ggtgtgagaa aggagccctt gggcaccacc 300
acacagagag gacagagatg actctcacan anancactct ctctctctca catatgtgtg 360
tggcagagat actctcacag cgccttgagg tggttaagcac tcgtgtgagt cacaatacgg 420
ctgtgtatct cacggcgtgg tgtgtagata gatgtgtgtt atctctccct gactctacac 480
acataatcta ctccacaca cacaacaat ctccgcagga acacacaaca aangggaaag 540
gaacgctgag caaagaccaa cgcaagaga gacaacaagg cagaagtcaa gcgagaca 598

<210> 63
 <211> 820
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (536)..(536)
 <223> a, c, g or t

<400> 63
 ccggccccggg ccggtgttgc gggcgccggg acactcccct cgcggggcg ccgtgcgtgt 60
 atgtgggtct cacacgattt cagggcctgg ctactgggat tgggtgccgc tcatggatgg 120
 atggccaatc tggggggata aagtgatgtg gcagttgtag taacttgccg gaacactgct 180
 cactctccgc agttcgtcgc tacgtgccta gtggttatag gacgtccctg ttgtcaacgc 240
 tcaggcgggt cgtggatcgc gttgggggta ccttcccagg catgtgctgt ggactattcg 300
 ggctttggag cgatggatcat tgcacacgcg taaaggagag ttccgcgcg tgtctgtgca 360
 ggcgtcctcg tacgctgcga gccgacttga tctggccctc ctcttgttg tagttgcccc 420
 tagttggccc cttttactac ctagtacgcg cctatatctt gaacttatct attcaattta 480
 ttacctttg gccaatgaa tttttggagg actcattgcc attgacacat tcactnagaa 540
 acggcgagaa aaggaagaac aaacagggaa caaaaataac acagcgtggg agaagacgaa 600
 cgagagggaa aaagcgggga gaccccgggg tgtgacacat tgggtggaaca cgcgggggac 660
 acacaaagtt ctcccagcaa aacacaacga cgcaaccaa ctaggagaca aaaaagaaca 720
 gaaaacaaag aaaaaaacg agaaaacgaa caaaaaaag acgaaagaac aaaaacaaaa 780
 aaaaaaagg gaaaaaaga aaaaagagaa agaagaagaa 820

<210> 64
 <211> 1305
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1021)..(1021)
 <223> a, c, g or t

<400> 64
 ccacgcgtc cgccggtgg accgcgtgc aggcacccg agggcgcggc aagatggagg 60
 tgacgggggt gtcggcacc acggtgaccg ttttcacag cagctccctc aacaccttcc 120
 gctccgagaa gcgatacagc cgcagcctca ccacgctga gttcaagtgt aaactggagt 180

40

tgctgggtggg	cagccctgct	tcctgcatgg	aactggagct	gtatggagtt	gacgacaagt	240
tctacagcaa	gctggatcaa	gaggatgcgc	tcctgggctc	ctaccctgta	gatgacggct	300
gccgcatecca	cgtcattgac	cacagtggcg	cccgccttgg	tgagtatgag	gacgtgtccc	360
gggtggagaa	gtacacgata	tcacaagaag	cctacgacca	gaggcaagac	acgggtccgct	420
ctttcctgaa	gcgcagcaag	ctcggccggt	acaacgagga	ggagcgggct	cagcaggagg	480
ccgaggccgc	ccagcgctg	gccgaggaga	agggcccaggc	cagctccatc	cccgtgggca	540
gccgctgtga	ggtctcacac	gatttcaggg	cctggctact	gggattgggtg	tccgctcatg	600
gatggatggc	caatctgggg	ggataaagtg	atgtggcagt	tgtagtaact	tgcgcgaaca	660
ctgctcactc	tccgcagttc	gtcgctacgt	gcctagtgg	tataggacgt	ccctgttgtc	720
aacgctcagg	cggttcgtgg	atcgcgttgg	gggtaccttc	ccaggcatgt	gctgtggact	780
attcgggctt	tggagcgatg	gtcattgcac	acgcgtaaag	gagagttcgc	gcgcgtgtct	840
gtgcaggcgt	cctcgtagcg	tgcgagccga	cttgatctgg	ccctcctcct	tgttgtagtt	900
gcccctagtt	ggcccctttt	actacctagt	acgcgcctat	atcttgaact	tatctattca	960
atattattac	ctttggccaa	ttgaattttt	ggaggactca	ttgccattga	cacattcact	1020
nagaaacggc	gagaaaagga	agaacaaaca	gggaacaaaa	ataacacagc	gtgggagaag	1080
acgaacgaga	gggaaaaagc	ggggagaccc	cgggggtgtga	cacattgggtg	gaacacgcgg	1140
gggacacaca	aagttctccc	agcaaaacac	aacgacgcaa	ccaaactagg	agacaaaaaa	1200
gaacagaaaa	caaagaaaaa	aaacgagaaa	acgaacaaaa	aaaagacgaa	agaacaaaaa	1260
caaaaacaaa	aaagggaaaa	aaagaaaaaa	gagaaagaag	aagaa		1305

<210> 65

<211> 759

<212> DNA

<213> Homo sapien

<400> 65

tcgcgcgcga	ggtacatgct	ggacaaccct	ctaactgacc	ctctcacctc	gtctttgtcg	60
actcccctac	agtgtcggcg	gtgaaccggg	gtggggcagc	acctcctagg	tccttccttc	120
tcctaacctc	ctccggggcc	gtgtcagctc	ctggctactg	tgactcaccg	acagtctgtc	180
ggcccagcaa	ctggacgaca	tctacttctc	gttgtcatag	cggcatgcgc	aagactgtga	240
aggacgccag	aacggtcgtc	ggctgagcac	aggggtccgt	ccacgactct	tgcctctcat	300
atctccccag	cgcgtccacg	cgactctccc	ttcccagagc	ctcctggggc	tgttccaccc	360
agcgatcagc	ctatgcgtat	gcgcggtgaa	ccccaaactg	ggaaggacag	cggcaaaaacg	420

41

tctgtcaatc ttttgccca atgaatctgc ccgctgtaga tgtgctggtc agtatcctcg 480
 cggggtccac gagtccccag tgccccaact ccttcaggac cagcgtgct tccttgggat 540
 cctaccccg cctccaggtga taagaagggg ggggttgccc ccgttaagaa aggccaaacc 600
 cccccccaa acaacgccgg cccggaaaaa aataaccaac gaacgtctgt ttttccttta 660
 acctgttcaa aaaaaaaca gaagaaaaa aagaaaaaa aaaaaaaaa agagccgtgg 720
 ggagaaaacc acggggggcac aaagggttta acccggggg 759

<210> 66

<211> 1450

<212> DNA

<213> Homo sapien

<400> 66

ctagaattac tacagaagca agtacgatgc agtgggatgg ggccgtgcct acactaaata 60
 tattgctggt tatcctggga cagctggaag ctgtgtcccc ttagccttac agaggcttgg 120
 tgggtgcatg ggccccatca ccttgacat ccaccttcat tgcctagaca gtgttagact 180
 atggagaaag acaactgaca ggatcgaga gccacggaga tcacatcaag gtgacgaggg 240
 gaagcagggg cattgatatg tctatagatt tggagcatta catatagttc gcaaggcatt 300
 tgggaagaaa aacctagtag ttgtattcca cctcgaaca cagcatatcc aagaccagt 360
 acaaaagagc gagaagatgc cataccacat aacggcatgg attcagggat cgacacgtcg 420
 taaacaccat tgtggcgata ctactatgg gacacttggg ggttcacaag agacaagcaa 480
 aaacaacaca cacagagcaa agaaagaaca agaagacaac aaaaagacg gagagagcga 540
 gggagaatac acacacaaag acgggggaca acaaaaagag gcagaggtac acacacgaca 600
 gtgggtgtac acaaaaacag gagacagaag aagcgaggcc acccaacaaa gaaatcaaca 660
 caccaacaag aagaagccac atccacgtcg gcgccgcgag agagggggga cggggaagac 720
 ggcaggcgag ggagaagaga agaaagagag acgaggcgcg gcgcagaaag agagagacga 780
 gaggacacgg cgagaaagga ggggaagcaga gaaaaaaga gaccaggag agaggagggc 840
 gcaagcgacg gggcagagcg gcaggaacac tcgaggagga gaaagggaga gcgagacaag 900
 gcagagagag aaggaggggg gagaaggaag aggaggacac caaggcgagg ggaaagagaa 960
 gagacaacgg aagaggagga agcacgagga agggcgacgg agagagcgag aaggccacga 1020
 gaagacagaa agggagagag gagaagacac agacgccaag aggaagcgca gaaggagcaa 1080
 agggagacgc aagagcaaga gcagggaag acagacgaga gagggaacgc agaaccacag 1140
 ggacgacaga aagagcagaa acgaaggaaa aagagcaggc acagcacgac agagacacaa 1200

42

acaagacaga aaaaaacgaa acgaaaagag aagagacgaa gcagcaacgc aaagaagaca 1260
 cagacagagg gaagaacgaa gagacgaagg aagaagagag cacacaaaaa gaggacaaag 1320
 atcaaaagag agaagaaaca caagaaaacg aagagacaga agaccagag acacgaaaca 1380
 agacgacaca gaaacacacg accaagaacg agcacagaag gaacagacac agaagagaca 1440
 cgaagaaaca 1450

<210> 67
 <211> 846
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (584)..(584)
 <223> a, c, g or t

<400> 67
 gagcggccgc ccgggcaggt acaaaaagaa ctcccacaac tcaacaacaa aaaaactgtt 60
 caaaaatggg caaaggtttg aatgtggttt tgttgagtc tatcttaaaa aaaaaaaaaa 120
 aaacaggcaa aggacttgaa tagacttctt tcttaagagg gtttgccatt gggcccaatt 180
 aagcttctaa caagatactg ggatatcact aatcattaag ggaaatgcta atcaaaacca 240
 caagatacca cctcatgccc attacaaggg ctactatcaa aagaacaaaa aataagtttg 300
 ggaagaatgt ggataaatgg gacactctgg tatactgggt gggggggaaa gtgaaatggg 360
 tatagccact ggtgggacaa tgttcagtc tcaaaaaatt aaactagaat tggccaaatg 420
 aaccagcaat ccacttctg ggtattatg cccaaaagaa aacagggtg gggcgaggg 480
 gcttacactt gtgatccag cacttgggga ggccaagtgg ggcggatcac aaggttcggg 540
 gatcgagacc atcctggcca catggtgaac cctgtctcta ctanaaatag aacatattag 600
 ctgggcgggg tggctggccc tgtatatccc agcacgtttg aagtcaggcg gtgggataac 660
 ctgaggtcgg gtcaagacca gctggccatg tggcaaaccc tgtctcttct aaataccaat 720
 tagtgggctg ggccctgcct gtaatccgct atcggaggct gggcggaat gctgactgaa 780
 gggagttgca gttggttaat ggtcgtacg gttttgacct atgagctggg actgcacgcc 840
 ggaagt 846

<210> 68
 <211> 326
 <212> DNA
 <213> Homo sapien

43

<400> 68
 tttttttttt tttttttttt tttttttttg ggaaaaaaa ataggcccggt gagtttttcc 60
 ttggatcact gggcactttt tgaaacctg cgggtgtgcca acccttcccc caagaggggg 120
 atttcagggg gggacccaat tcctcagggg tggaagtgac ctgggaccac acagagccgg 180
 gactctacag acccctccct cattagtggc aggaatacag cttgggttac ccagtgtca 240
 tagcctgtcg cgtgtgtgaa atggttactc gcctcaaata cacacaacat ccgagcaaat 300
 taacaaggca caaaaaacca aaccga 326

<210> 69
 <211> 886
 <212> DNA
 <213> Homo sapien

<400> 69
 gctggccgcc cgggctggta cgtgccaagc tttcctcgtc gagcgacccc gagagccctg 60
 gggagcgggtg ggcttgccgg cccgatcgca ctcatattatc ccggagacag gggatatgagg 120
 ctctctcgtg cgatgtatgt gggttgtagc agcatgcctc atgcgatcac ggcagcatga 180
 ggaagacgtt cccctgcatg ccacctgcgt ctatggtccg acggttgaag cgttgctgac 240
 taggcaggtg acatgacggg taggcggtc gcgagttcca aggcagcggc gcatggcgggt 300
 gggtccttcg tacgtgtgtt cgtccgtgcc ttgccgctag ttgctcttc ccatctcgca 360
 tcgagtcgaa ctgtttctgc tgtggaatta tgacacgcgc tttgagcgt agggcgacgg 420
 ttctacgtgc ttgaactcgt tgggtgttct tttgggtcga gtgtctcatt ctgcgggagt 480
 gactgggggtg cgccaggggc tgcttaccat caggggccccg tgcagactct aacgccataa 540
 agcccgtcaa tcttctgag cattgacctt tgggcccggc ctgatataac tttagggtggg 600
 attccctgag tcgctcgagt taaccaaagt catatgggtg ggtagatcga tgggattcaa 660
 ttacgccttt ggcttatcct cgcgttggtc cgttttgaca aagttcgttt ccttttcccg 720
 ggagttccac ccagaatttt tccccatcta actaaacata ttctgaacg gacgcgcaaa 780
 acaaaagggg aaccgcacta atcaatacga accactcata catccagccg tgaaacacgc 840
 agaagccatc gccaacgaga gaccaccaca caaacggcac cctaga 886

<210> 70
 <211> 747
 <212> DNA
 <213> Homo sapien

<400> 70
 gataatcata tggcaatggg cctctaatagc atgctcgagc ggcgagtgat gatggattgg 60

44

ctcgcggacg tggtaacct gattcaaata taattccatg agaagctgga ctaaggacat 120
 atattcattc attcaatatt catgtgtgtg tgtgttagag acaggggtcg agcggcggat 180
 ggtgagatcg gctgttcgca gcaggggacg gtcgggatgg acaaagacgc ggactccaga 240
 tctggggagc cctgcagacc tagaccgcat cttgcctgac tcagttccct gaagctcttg 300
 tcgatctact ggccgtgaag gcgtaacaaat tctaggggtca gctgtatgca ctaagctcaa 360
 aagtcaagtg agattgcttc ctccctctcg caagccgaag cccaaaagtt cttgagcaac 420
 gatgtagtgt acatcgcagg agaggagcga taatacagtc agttgttatg cacttaatat 480
 agggaacgat ttctcgatgt gctgatcacg acaaacact ttaagtttga ccaggaattc 540
 taacctgtcg tctggtagta tatagtgcac taaaagtgtg ggtgtcaaca tattactaaa 600
 gcaaagctta tatttcaatt aataaggggg agttaggcgc ctttataagt ctagcaaaag 660
 agtaaaaaca acacacagag agcctgggcg acacgggaga ccatacgggt cccgggggga 720
 agtggttatgc cggacaacca caaaaag 747

<210> 71
 <211> 1374
 <212> DNA
 <213> Homo sapien

<400> 71
 tgctgtgtgg atgatgtgct gtgttttgcg tatttgtggc tgctcttcct gcttccatct 60
 gcatcactag ccatccgact tgcgctctgt gtctgttctg gctgctgtgc gagctggat 120
 catgctcact caaatgtgct gtgtaatact gtgttatcca catgaatcat gtgggataac 180
 tgcattgtgaa atgaacatcg ttgatgcaaa atgtgccatg caaaatgtgc catgtgaacc 240
 tgtaaacaat ggccggcgat ttgcgttggc tgtaagtatg tccagcaacg agacatactg 300
 agtccagagg acaggggaga agtagaagac acaggagaag agagaggag aagagaagaa 360
 aggacaacaa acggagacca gcgcggcagc aacaagcacg cgacagaggc gagagcagag 420
 acaggagaag agagaccagc gacggcagcg acagagcacc gaacaacaaa gagggcgatg 480
 gcgggtcgca gcaacagaag aagggtgata gagatatgac ggaggaggag aatgagttag 540
 aagaggtgat gtctattaga gagcccaaag acgaagaaga aaatgagaag agagcgaccg 600
 agggatagaa gaagaagaaa aggcgcaggg acgagggaga gggagcgcga aggggaagag 660
 gaggcgcaag cggcgccagg acaggccgag agcgggaagag gcgggaagag cgacagagag 720
 gcgagggcac acaagaaggc aagcgacacc agaagagaca ccgcgcagag caagggggag 780
 gacctcagaga gagaggggag aaggagagag ggggacgcag accccaggac gcggggccag 840

45

gccaggagac acggaaggca gagagaggag ggaggaggca agaagagaga gcggaaccagg 900
 gggccacgca gcggagacga ggagagacga gcgcaggcac agcggcaggc gcgggggcgg 960
 ggaggagagc gagcgagcgg gcgagcgaga aggagcgtca ccaaaggagg gaggagcaga 1020
 ccgcacagcc gaagagcgcg gcgccaggaa acggaagaga gagagagagg agaccgagga 1080
 cgaacacgag aaaaggggag gacgagacgg cggaacagag cgcgcgagc acggccgccg 1140
 agcaccggc aagcaggcac cacgagacag accgagcggga agcaggcgcg cgcaagaccg 1200
 aacaggccaa cgagcgaagc acgagcgaag cgacaagcgc accaagatgc gaaccaggca 1260
 gccgacgaag aagggcaaga ccacaagaga agacgacccg ggagagagga gagaggacag 1320
 ccagaggcag cacacacgaa cagctgaggg ccgaagccga gccacagacg gcac 1374

<210> 72
 <211> 578
 <212> DNA
 <213> Homo sapien

<400> 72
 cacaacatac gagcatacga gcatctcccc aaatcccggc cgccacgggc caggacaagt 60
 gcccactcc ttgaatcacc gcaatgatct ttttccgttc accggccccg tcagggcacc 120
 ccgagatgtc tccaacctcg caccaatagt taacagcggc cgaagcacc cctcaggggc 180
 cccttgacat acatcttggt cacaacagca gcccagagaa gcatgtggct gggggagcat 240
 aacacagacc atggaacttt ctgtctcgga gagaacacc ggggtgatcc cttctgcaaa 300
 tagctggcgg atcaccgaag tgcacagcgt agagtcatcc ccaccgttcg agaggactct 360
 cacatccagt gattgaacac acttactctt tatcacacca ggtggtgaga gctgtctaga 420
 tggacctcgc agactagatc atagaccccc ttcgaccgtt ggatgccggg ggtcatggga 480
 gggccattgg ggccagtccc gagccgacat tgctgcgga tggtttaact cagaaccccc 540
 tgaacgtaag gccgaagaac ctacgagaat cccctgt 578

<210> 73
 <211> 700
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc feature
 <222> (510)..(510)
 <223> a, c, g or t

<400> 73
 gcggccgcgc gggcaggtag ccaacagctc attgagaacg ggccaggatg acaatggcgg 60

ctttgtggaa tagaaaggcg ggaaagggtg ggaaaagatt gagaaatcgg atggttgccg 120
 tgtctgtgtg gaaagaagta gacatgggag acttttcatt ttgttctaca ctaagaaaaa 180
 ttctctgcc ttgggacct gttgatctgt gaccttacct ccaaccctgt gctctctgaa 240
 acatgtgcgg tgtccactca gggttaaatg gattaagggc agtgcaagat gtgctttggt 300
 aaacagatgc ttgaaggcag catgctcggt aagagtcac accaatccct aatctcaagt 360
 aatcagggac acaaacactg cggaaggccg cagggtcctc tgcctaggaa aaccagagac 420
 ctttgttcac ttgtttatct gctgaccttc cctccactat tgtcccatga ccctgcaaaa 480
 taccctctg tgaggaaaca cccaagaatn atctaaaaa aaaaaaaaaa aaaaaaaaaa 540
 aaggcttggg ggttaccagt ggccaatagc gtgttccccg ggggttgaat tggtttcccg 600
 cctcaactcc ccattctta gacaacaaaa gtccgcaaag agatcatcaa tagtcaatca 660
 acctcattac gaaacacaag acaatgaaat aaaataacaa 700

<210> 74

<211> 815

<212> DNA

<213> Homo sapien

<400> 74

ccgcccgggc aggttgtaat ccagctact tgggaggctg aggcagagaa ttgcttgaac 60
 ccgggaggca gaggttgagc tgagtcgaga tcgtaccact gcactccagc cagggaaca 120
 gaaggagact ccatctcaaa aaaaagaaaa aaaggttaagg ccgaactcag ttgctcacac 180
 ttgtaatctc agcacttcgg gaggaggctg aggcaggcag attgcttgcg cttaggagtt 240
 caggactgaa ctaggcaaca tggagaaacc atgtctctac aaaatataaa aaaattagct 300
 ggacatggtg tcttgacact gtagtcccag ctactcagga ggctgagctg ggagtatcac 360
 ttgagcccag gaagtgcaga ttgcagtagc caagatcatg ccactgcact ccagcctggg 420
 aaacatagtg agatcctgtc tcaaaaataa taataataaa ataggccgag cgcggtggct 480
 cagcctgta atcccagcac tttgggaggc caaggcgggt ggatcacgag gtcaggagat 540
 caagaccatc ctggctaaca cggtgaaacc ccatctctac taaaaataca aaaaattagc 600
 ccggtgtggt ggtgggccc ttagtccca gctactaggg aggcggaggc aggagaatgg 660
 cgtgaacccg gggagggtga gcttgagtg agccgagatt gcaccactgc actccagcct 720
 gggatacag cgagactcca tccccaaaa aaaaaaaaaa aaagctgggg ttacctggcc 780
 aaagggttcc ggttgaatt ggttccgcc caatc 815

<210> 75
 <211> 880
 <212> DNA
 <213> Homo sapien

<400> 75
 cgagcggcgc ccgggcaggt acccctatcg tatctaggaa gtaactagct tgcttttcat 60
 tttacaggct cataggcaga agggacttgc cttgtctcaa atgagacttt ggactatgga 120
 cttttggggt aatgctgaaa tgagttaaga ctttggggga ctgttggtgaa ggcacgattg 180
 gttttgaaat gtgaggacat gagatttaga gggggccagg ggcagaatga tatggttggg 240
 ctgtgtcccc acccaaactc caacttgaat tgtatctccc agaattccta cacgttgtgg 300
 gagggaccca ggagaggtaa ttgaatcatg ggggctggtc tttaccatgc cattcttata 360
 atagtgaata cgtctcatga gatctgatgg gattatcagg gacttccgct tttgcttctt 420
 cctcgttttc tcttgctgcc accatgtaag aagtgccttt cgcctctcac catgattctg 480
 aggcctcccc agccatgtgg aactgtaagt ccaactaaac ctttttttct ctccagtctc 540
 aggtatgtct ttatcagcag catgaaaata tactaataca tatttcatgt taatgggctg 600
 ggagtattta gtatttggtta agatgggttt gaattatttg tcccccttc aaaaactcat 660
 gtttgaccac attacttccc caaaaatgac tgtatttgag aggtgggggc ctttaaggat 720
 gttgattggg gtcacgggga tctgctttct gactggcttc acattggggg cttgcctgcg 780
 gatggattaa tggctatgcg gattggagat gttggctcct caacagagga cacagagctg 840
 gagaactccc aacccttcca tttatgccgc atacttccac 880

<210> 76
 <211> 1666
 <212> DNA
 <213> Homo sapien

<400> 76
 atggctgaaa agggccaaca tagagctcag gctatggctt cagaggggtg agggcccaag 60
 ccttggcagc ttccacatgg tgctgagcct gcagggtgcac agaagtcaag aattgaggtt 120
 tgggaacctc catctagatt tcagaagatg tatggaaatg cctggatgcc caggcaaaag 180
 ttggcatcag ggtcacagcc cttatggaaa acttctgccg gggcactgtg gaagcaaatt 240
 gtggggtcag agccacaca cagagtcctt aatggggcac tgcctagtag agctgtgaga 300
 agagggtcac cttcctccaa accccagaat ggtagacctt ccaacagctt gcaccgtgag 360
 cctggaaaag tcacagacac tcagtgccag cccatgaagg cagccaggaa agaggttgta 420
 ccctgcaaag ccacagggtt ggtggagctg cccaagacca tgggaagcaa actctttcat 480

48

cagtgtgacc tggatgtgag acctggagtc aaaggagatc attttggagc tttaaaattt 540
 gaagctctgc tggatttcag acttatgtgg gccctgtcac ccctttgttt tggccaattt 600
 atcctatttg aaatggctgt atttacccaa tacctgtacc tcccttgat ctgggaagta 660
 actagcttgc ttttcatttt acaggctcat aggcagaagg gacttgcctt gtctcaaatg 720
 agactttgga ctatggactt ttgggttaat gctgaaatga gttaagactt tgggggactg 780
 ttgtgaaggc acgattgggt ttgaaatgtg aggacatgag atttagaggg gccaggggc 840
 agaatgatat ggttgggctg tgtccccacc caaatctcaa cttgaattgt atctcccaga 900
 attcctacac gttgtgggag ggaccagga gaggttaattg aatcatgggg gctggtcttt 960
 accatgccat tcttataata gtgaatacgt ctcattgagat ctgatgggat tatcaggggac 1020
 ttccgctttt gcttcttctt cgttttctct tgctgccacc atgtaagaag tgcctttcgc 1080
 ctctcaccat gattctgagg cctccccagc catgtggaac tgtaagtcca actaaacctt 1140
 tttttctctc cagtctcagg tatgtcttta tcagcagcat gaaaatatac taatacatat 1200
 ttcattgtaa tgggctggga gatttagtat tgttaagatg gtttgaatat ttgtccctt 1260
 caaaactcat gttgaaaatt aattcccaa atgacagtat tgagaggtgg gccctttaag 1320
 aagtgattgg gtcattgagg atctgctttc atgaatgat tagaaatggg gtcttgctgt 1380
 gaatggatta atggcttatg ctggaattga gactgttggc ttcataagaa gaggaagaga 1440
 gatctgagct agcatcctca gccccttgc catatgatgc cctgcattac ttccagactc 1500
 tgcagagaga ccttaccagc aagaaagccc tcaccaaagt cagcccctca accttgact 1560
 tctgagcctc tataattcta agaaataaaa tctgttctt tataaaaaaa aaaaagaaaa 1620
 aaaaaagaa aaaaagaaaa accgagaaac tcgagggggc ccgtaa 1666

<210> 77

<211> 87

<212> DNA

<213> Homo sapien

<400> 77

ggatgttaat cactatagcg atggtgctct agatgctctc gagcggcgca tgtgatggat 60

ccgctccttt acagccctgc gcctgaa 87

<210> 78

<211> 458

<212> DNA

<213> Homo sapien

<400> 78

gatgatgac atataggcga atgggtctct agatgcatgc tcgagcggcg cagtgtgat 60

gatgattggt tgaccacagg agtccgagac cagccgggggt aacatggcgg gaccccaatc 120
 tctaccaaaa aaaaaaaaaa tacaaaagtt gtcgggggtgt ggtgtgcttg cctgtagtcc 180
 caagtcccag ctactctact tgggaggctg aggcagaagg gattcaccgt gagcccagga 240
 gggccagggc ttgcagttag ccccgtagt ggtgccactg tgcacttgac cttggggggca 300
 acagaagtga gaattgagac ccctggttca aaaaaaaaaa aaaaaaaaaa aaaaaaaggc 360
 ggttgggggt tcttcagggg gctcatgggt gtgttccgtg ggtgtgaaat tgtgtttctc 420
 ccggctccaa aattttctca caaaaatatt gaaaaaaa 458

<210> 79
 <211> 905
 <212> DNA
 <213> Homo sapien

<400> 79
 actatttcaa caagcttttt catgtaacta atctgcggaa ggtagaaagg ggaaaactgt 60
 tgggtgctaa aatgacaact ggttcaagggt acaatggcga atatttttat ttctgcaact 120
 tttcttagag gttggaaact ggactgggca ggaagattcc tttttgtaag attagtctcc 180
 agttttcatc aagcagttta gtgggggtatt ttaggccag ttcctctctc acagtcccca 240
 aaggctcttct gttaacttta aatccgcaaa gagagagatc tctgccaagc agcaactgca 300
 agagcatgtg ggtcaatgtt accagcagac actcaaagcc ccttcccttt acttcaacac 360
 cgctttataa attatcttag agacgttgtc aggttggtat tagagggtgag tgggtcatgac 420
 ttcacgattt ctcatctttc tgaatgcata gtggctggga gtgggtggctc atgcctgtaa 480
 tcccagcgggt ttgggaggcc gaggtgggca gattgtttga ccacaggagt tcgagaccag 540
 ccggggtaac atggcaggac cccaatctct accaaaaaaa aaaaaaatac aaaagttgtc 600
 ggggtgtggt gtgcttgctt gtagtcccaa gtcccagcta ctctacttgg gaggctgagg 660
 cagaagggat tcaccgtgag ccagaggagg ccagggttg cagtgagccc cgtgattggt 720
 gccactgtgc acttgacctt gggggcaaca gaagtgagaa ttgagacccc tggttcaaaa 780
 aaaaaaaaaa aaaaaaaaaa aaaaggcgggt tgggggttct tcagggggct catgggtgtg 840
 ttccgtgggt gtgaaattgt gtttctcccg gtcctaaaat ttctccacaa aaatattgaa 900
 aaaaaa 905

<210> 80
 <211> 1381
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (282)..(282)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (375)..(375)
 <223> a, c, g or t

<400> 80
 cgagcgggcg gccggggcag gtacttctac tgcccaagat gctaccattt accgtggaga 60
 ggtgtgcttc tgtagatttc ttgagtgate ctggaaatgt cccattcgat ggacggacga 120
 tgccgcgaat gcgtccatgc gctggtgaac tagagtgaag ggcattgagcc acttgccgtg 180
 gagggcatga tcagaacgac ttgcggagtgc caatctgatt cgtggcctgt tgccccgagt 240
 tctcgtacgt ggaattagct gaccaccgtg acaaggccga cntctctagt ggccagtga 300
 actgtgtggt aaagatggga tatatgtacc ctgctgtgta gcggtgggac atatgatatg 360
 tcgctggggg aatcnatgct gtactcgtga ctgacctca tcgaaatgta ctgtcgtaag 420
 tagttgagtg ggcacctccc aagataggat agaatgcctg ggtttgatgc aaggcatagc 480
 taaaggagga cactgcgcac tggggaccgc aggggtgggt ggagcgcaat catttcctag 540
 gcccgttccg aagaacgtat tgaattgatg cgttgtccg aggtaaggca cccttttagat 600
 tagacatagc tggtagggca ataactatct cgatgcaatg ctgtacgata tacgcctttt 660
 ggacaactcg tctatagtgt ataaccaatt gggtaattgg ccgaattaga accaggtaca 720
 catggatcct tcatccgcac gttccgggtc ggccagaaga acccgcctcg ttggcgcttg 780
 ggggcctttt cgaatcacia ccacgttgcc cccgaatctg gaaataaaaa gttggcgtgg 840
 ggacaagtaa actttaagta agcatccttt tcccaccaa aaacgtacac cttttacttg 900
 ggtgttagaa tgtagcccca aaggaaatcc tggggtcaaa gggaaatggtt aaaggaaggg 960
 gccaacatcg atggaattga gcgaccggtt ggtacgcttt ggggggtaaa agttagacag 1020
 acacagttcc ccgaaaggca cttttaagca ggaggtactt ggaactttgt gaacccatgt 1080
 aaaaggggtt tttagtgtgg gcgggagtta gctcccata ggggaaagt gggtttgtga 1140
 accctcccaa cggtggcgcc gaggtgtaaa atccggttcc ctacaatttt tggcgctatg 1200
 aagttgggag ttaagtatag ggaaaaccac ttatcacagg tatgggcacc agatagagat 1260
 agagacaaat ctgtgtgggg ggggaaataa ccaggtgggg ggtcagcaaa gggggggggg 1320
 caacccccgg ggggggtaga aagagggggg tatatgcccg ggcacaaggg gatcccgaga 1380

g

1381

<210> 81
 <211> 668
 <212> DNA
 <213> Homo sapien

<400> 81
 gccgcccgagggtaccca acagctcatt gagaacgggc caggatgaca atggcggctt 60
 tgtggaatag aaaggcggga aagggtggga aaagattgag aaatcggatg gttgccgtgt 120
 ctgtgtggaa agaagtagac atgggaggct tttcattttg ttctacacta agaaaaattc 180
 ctctgccttg ggatcctggt gatctgtgac cttaccccca accctgtgct ctctgaaaca 240
 tgtgcggtgt ccactcaggg tttaatggat taagggcagt gcaagatgtg ctttgttaaa 300
 cagatgcttg aaggcagcat gtcgttaag agtcatcacc aatcccta atctcaagtaat 360
 cagggaacaca aacactgcgg aaggccgcag ggtcctctgc ctaggaaaac cagagacctt 420
 tggtcacttg tttatctgct gaccttcctt ccactattgt cccatgacct tgccaaatac 480
 ccctctgtga gaaacaccca agaattatct aaaaagaaaa aagaagaaaa aaaaaagaaa 540
 aaaggcgggg ggtaaacctg gggcagaagc ggtgccctgg ggggaattgg gttttcccgt 600
 ccccatctcc cccactctg cgcgcaaaaa cggtaagcaa agagaacagg agcagagaga 660
 caggaaag 668

<210> 82
 <211> 7626
 <212> DNA
 <213> Homo sapien

<400> 82
 gttgacccgc ggcgttcacg ggaactgttc gctttagtgc cggcgccatg gggtcggagc 60
 tgatcgggcg cctagccccg cgcctgggco tcgccgagcc cgacatgctg aggaaagcag 120
 aggagtactt gcgcctgtcc cgggtgaagt gtgtcggcct ctccgcacgc accacggaga 180
 ccagcagtgc agtcatgtgc ctggaccttg cagcttcctg gatgaagtgc cccttggaca 240
 gggcttattt aattaaactt tctggtttga acaaggagac gtatcagagc tgtcttaaat 300
 cttttgagtg tttactgggc ctgaattcaa atattggaat aagagacctt gctgtacagt 360
 ttagctgtat agaagcagtg aacatggctt caaagatact aaaaagctat gagtccagtc 420
 ttccccagac acagcaagtg gatcttgact tatccaggcc acttttctact tctgctgcac 480
 tgctttcagc atgcaagtag gtatttcatt aaacattcag aaaagttacc aatttacaag 540

tgggtttttc atccccaagg aatacttcta acttagttga tatcaattca gagcatattt 600
tcccctagaa ataattattag gaattattggc caagtgacta tattcccagt ttatoccata 660
atgtagctaa caacttggaa ctagtggtgc cagaattcca ctagcaaata gcagctgtat 720
atatatgctg ggaattctga tttcagctcg ccttttgtaa gagatgatat ctgtcattaa 780
aacagtcttc acatgagatt tttctgctca ttttttttaa aaagtactgg ttgggccagg 840
cgtgggtggc cccgcctgta atcccaacac tgggaggcag aggcaggagg actgcttgag 900
gcaaggagtt caagactagc ctagacagca taataagacc ccaatctctt aagaaaaaaa 960
aaaaaaatta gctgggtgtc agcacatgcc tccagtcctg gcttctcagc tactcgggag 1020
gctgaagctg aaggctcact ggagcctagg agttcttggg tatagtgagc tatggtcacg 1080
ctactacact gcagcctagg caacacagca acactgtctc tttttttttt tttttttttt 1140
tttttttttt tttctacaat tctttttttt tttttttttt ttaatttatt tttttattga 1200
taattcttgg gtgtttctca cagaggggga tttggcaggg tcatgggaca atagtggagg 1260
gaaggtcagc agataaacia gtgaacaaag gtctctgggt ttcctaggca gaggacctg 1320
cggccttccg cagtgtttgt gtccctgatt acttgagatt agggattggg gatgaactct 1380
aacgagcatg ctgccttcaa gcatctgttt aacaaagcac atcttgacc gcccttaatc 1440
catttaaccc tgagtggaca cagcacatgt ttcagagagc acaggggttg gggtaaagtc 1500
acagatcaac aggatcccaa ggcagaggaa tttttcttag tgcagaacia aatgaaaagt 1560
ctcccatgtc tacttctttc tacacagaca cggcaaccat ccgatttctc aatcttttcc 1620
ccacctttcc cgcctttcta ttccacaaag ccgcattgt catcctggcc cgttctcaat 1680
gagctgttgg gcacacctcc cagacggggg ggtggctggg cagaggggct cctcacttcc 1740
cagtaggggc ggccgggcag aggcgcccct cacctcccag acggggcggc tggccgggcg 1800
ggggggctga ccccccccc accggtcagg ttagtctcga actcctgacc tcatgatctg 1860
ctcacctagg cctoccaaag tgctgggatt ataggcatga gccactgcac ctggccagt 1920
gataagcttt ttgatgtgct gctggattca gtttgccagt atgtgattgt ggatttttac 1980
atcgatgttc atcagggata tgccccgcc agagaggagg aatctagaga ggcagtctgg 2040
ctacagcagc tttgccaagc tgcaaggggc tctgcccagt ccaaaattcc cagcggggtt 2100
gtttacattg tgaggggaaa agcacctact caagcctcag ttatggcagt tgccctccc 2160
cccaccaagc tccagggtcc caggtgtcct tcagactgct gtgctggcaa tgagaatttc 2220
aagccagtgg atcttagctt gctgggctcc acaggggttg gatccactga gctagaccac 2280
ttagctccct ggttcagcc cctttccag attatggccc taagtgaacc agagtatagt 2340

tattttctcca	ttttatttga	cagcaccctg	gagacaacat	ttgacagcac	tgtgacaaca	2400
gaagttaatg	gaaggaccat	acccaacttg	acaagtcgac	ccacccccat	gacctggagg	2460
ttggggccagg	catgtccgcg	acttcaggcg	ggagatgctc	cctccctggg	tgctggctat	2520
cctcgcagtg	gtaccagtgc	attcatccac	acagaccctc	cgaggttcat	gtataccacg	2580
cctctccgctc	gagctgctgt	ctctaggctg	ggaaacatgt	cacagattga	catgagtggag	2640
aaagcaagca	gtgacctgga	catgtcttct	gaggtcgatg	tgggtggata	tatgagtgat	2700
ggtgatatcc	ttgggaaaag	tctcaggact	gatgacatca	acagtgggta	catgacagat	2760
ggaggactta	acctatatac	tagaagtctg	aaccgaatac	cagacacagc	aacttcccgg	2820
gacatcatcc	agagaggggt	tcacgatgtg	acagtggatg	cagacagttc	cttgaagttt	2880
ctcaccgaga	tagagctggg	gatgatacca	cctgtgcaga	gtgaaaattc	caccagtcac	2940
gagaagcctg	cccagcaggt	ttgtcaaaga	tcagatagtt	gtagatatgc	ggccctcagc	3000
aatgttaaaa	gaacagaaat	tatacaaac	tatctctcag	accacagtgc	aatcaaaacta	3060
gaactcagga	ttaagaatct	cactcaaagc	cgctcaacta	catggaaact	gaacaacctg	3120
ctcctgaatg	actactgggt	acataacgaa	atgaaggcag	aaataaagat	gttctttgaa	3180
accaacgaga	acaaagacac	aacataccag	aatctctggg	acgcattcaa	agcagtgtgt	3240
agagggaaat	ttatagcact	aatgcccac	aagagaaagc	aggaaagatc	caaaattgac	3300
accctaacat	cacaattaaa	agaactagaa	aagcaagagc	aaacacattc	aaaagctagc	3360
agaaggcaag	aaataactaa	aatcagagca	gaactgaagg	aaatagagac	acaaaaaaccc	3420
cttcaaaaaa	tcaatgaatc	caggagctgg	ttttttgaaa	ggatcaacaa	aattgataga	3480
ccgctagcaa	gactaataaa	gaaaaaaaga	gagaagaatc	aaatagacac	aataaaaaat	3540
gataaagggg	atatcaccac	cgatcccaca	gaaatacaaa	ctaccatcag	agaatactac	3600
aaacacctct	acgcaaataa	actagaaaat	ctagaagaaa	tggataaatt	cctcgacaca	3660
gacactctcc	caagactaaa	ccaggaagaa	gttgaatctc	tgaatagacc	aataacagga	3720
gctgaaattg	tggcaataat	caatagctta	ccaacaaaaa	agagtccagg	accagatgga	3780
ttcacagccg	aattctacca	gaggtacaag	gaggaactgg	taccattcct	tctgaaacta	3840
ttccaatcaa	tagaaaaaga	gggaatcctc	cctaactcat	tttatgaggc	cagcatcatt	3900
ctgataccaa	agccaggcag	agacacaacc	aaaaaagaga	atttttagacc	aatatccttg	3960
atgaacattg	atgcaaaaaat	cctcaataaa	atactggcaa	aacgaatcca	gcagcacatc	4020
aaaaagctta	tccaccatga	tcaagtgggc	ttcatccctg	ggatgcaagg	ctggttcaat	4080

atacgcaa at caataaatgt aatccagcat ataaacagag ccaaagacaa aaaccacatg 4140
attatctcaa tagatgcaga aaaagccttt gacaaaattc aacaaccctt catgctaaaa 4200
actctcaata aattaggtat tgatgggacg tatttcacaaa taataagagc tatctatgac 4260
aaaccacag ccaatatcat actgaatggg caaaaactgg aagcattccc tttgaaaact 4320
ggcacaagac agggatgccc tctctcaccg ctcctattca acatagtgtt ggaagttctg 4380
gccagggcaa tcaggcagga gaaggaaata aagggtattc aattaggaaa agaggaagtc 4440
aaattgtccc tgtttgacaga cgacatgatt gtttatctag aaaaccccat cgtctcagcc 4500
caaatctcc ttaagctgat aagcaacttc agcaaagtct caggatacaa aatcaatgta 4560
caaaaatcac aagcattctt atacaccaac aacagacaaa cagagagcca aatcatgagt 4620
gaactcccat tcacaattgc ttcaaagaga ataaaatacc taggaatcca acttacaagg 4680
gatgtgaagg acctcttcaa ggagaactac aaaccactgc tcaaggaaat aaaagaggac 4740
acaaacaaat ggaagaacat tccatgctca tgggtaggaa gaatcaatat cgtgaaaatg 4800
gccatactgc ccaaggtaat ttacagattc aatgccatcc ccatcaagct accaatgact 4860
ttcttcacag aattggaaaa aactacttta aagttcatat ggaacacaaa aagagcccgc 4920
attgccaagt caatcctaag ccaaaagaac aaagctggag gcatcacact acctgacttc 4980
aaactatact acaaggctac agtaacacaaa acagcatggg actggtacca aaacagagat 5040
atagatcaat ggaacagaac agagccctca gaaataatgc cacatatcta caactatctg 5100
atctttgaca aacctgagaa aaacaagcaa tggggaaagg attccctatt taataaatgg 5160
tgctgggaaa actggctagc catatgtaga aagctgaaac tggatccctt ccttacacct 5220
tatacaaaaa tcaattcaag atggattaaa gatttaaacg ttagacctaa aaccataaaa 5280
accctagaag aaaacctagg cattaccatt caggacatag gcgtgggcaa ggacttcatg 5340
tccaaaacac caaagcaat ggcaacaaaa gccaaaattg acaaatggga tctaattaaa 5400
ctaaagagct tctgcacagc aaaagaaact accatcagag tgaacaggca acctacaaca 5460
tgggagaaaa ttttcgcaac ctactcatct gacaaagggc taatatccag aatctacaat 5520
gaactcaaac aaatttacia gaaaaaaaca aacaacccca tcaaaaagtg ggccaaggac 5580
atgaacagac acttctcaaa agaagacatt tatgcagcca aaaaacacat gaaaaaatgc 5640
tcatcatcac tggccatcag agaaatgcaa atcaaaacca ctatgagata tcatctcaca 5700
ccagttagaa tggcaatcat taaaaagtca ggaaacaaca gcaaaaagaa caaagctgga 5760
ggaatcatgc cagctgactt caaactatac tacaaggcta tgggaacaaa aacagcatgg 5820
gacatggatg aagctggaaa ccatcattct cagcaaactg tcgcaaggac aaaaaaccaa 5880

acgccgcgtg ttctcactca taggtgggaa ttgaacaatg agaacacttg gacacaggaa 5940
ggggaacatc acacactggg ccttgatcatg cgtttcgggg ctagggaagg gatagcatta 6000
ggagaaatac ctaatgtagg cacactcaca ctctcactg gctatggggg atgccagctg 6060
ccatgctgca aggacactca ggcagcctat ggagaaaccc acgtggtgcg gagtggaggc 6120
cttctgccaa cagccagctg ggaactgagg cctgctgaca gtcacacggg gaccagcgat 6180
gatccaggcg tctcggtcgt tagcgggtat cctgggggct gtctccctga ccacgacccc 6240
ccagtggggg ttcttttcga ggggtcccgc cctcgagctg gctctttgat aaagggcgga 6300
ggaacggggc tggctgcttc ccgagtcctc aggtcccgcg agcggcgggc gtgttgccgg 6360
tatggggtgc ggcgccagca ggaaggtggt cccggggcca ccagcgctgg cttgggcca 6420
gcacgaaggt caaaaccaag ccggcgctcg aggcggcggg cctggggccg aggcggcggc 6480
ccaggcgggc cagaggatac aggtggctcg ctccgagcc aagttcgacc cccgggtcct 6540
tgccagtgcc cagtacaatt tctctttgac atctctgaac agggagttca gaggatggga 6600
aaaaagagag caggagcagc agcaacaag ggaaggaatt cctatcttcg gagatatgac 6660
atcaaagctc ttattgggac aggcagtttc agcagggttg tcagggtaga gcagaagacc 6720
accaagaaac cttttgcaat aaaagtgatg gaaaccagag agagggaagg tagagaagcg 6780
tgcggtgtctg agctgagcgt cctgcggcgg gttagccatc gttacattgt ccagctcatg 6840
gagatctttg agactgagga tcaagtttac atggtaatgg agctggctac cggaggggag 6900
ctctttgatc gactcattgc tcagggatcc ttacagagc gggatgccgt caggatcctc 6960
cagatgggtg ctgatgggat taggtatttg catgcgctgc agataactca taggaatcta 7020
aagcctgaaa acctcttata ctatcatcca ggtgaagagt cgaaaatttt aattacagat 7080
tttggtttgg cactactccg gaaaaaagt ggtgactgga caatgaagac actctgtggg 7140
accccagagt acatagctcc tgaggttttg ctaaggaagc cttataccag tgcagtggac 7200
atgtgggctc ttggtgtgat cacatatgct ttacttagcg gattcctgcc ttttgatgat 7260
gaaagccaga caaggcttta caggaagatt ctgaaaggca aatataatta tacaggagag 7320
ccttggccaa gcatttccca cttggcgaag gactttatag acaaactact gattttggag 7380
gctggtcatc gcatgtcagc tggccaggcc ctggaccatc cctgggtgat caccatggct 7440
gcagggtctt ccatgaagaa tctccagagg gccatatccc gaaacctcat gcagagggcc 7500
tctcccact ctccagagtc tggatctgca cagtcttcta agtcacatta ttctcaciaa 7560
tccaggcata tgtggagcaa gagaaactta aggatagtag aatcgccact gtctgcgctt 7620

56

ttgtaa

7626

<210> 83
 <211> 384
 <212> DNA
 <213> Homo sapien

<400> 83
 taactcccat ttgccaatat ggaaagatga gcaaggccag tcaactgtggc tcatgcctgt 60
 agtctcagca ctttgggagg cagaggcagg aggatcgctt gaaccgagga gtctgagggt 120
 gcagtgagtt gtgatagtgc cactgcactc cagcctgggg tgacagactg agagaaagaa 180
 aggaagaaag gaagaaagaa agaaagagag agagagagag agagaaagag aaggaaagaa 240
 agaaagaaga aagaaaggaa agaaagaaga aagaaaaaga aagaaaaaaa aaaaaaaggc 300
 tgggcgtaac tcagtggctc atagcgtggt cccgtgggtg gaaatgtgtt attccgctca 360
 caattctcca cacaacattt caac 384

<210> 84
 <211> 482
 <212> DNA
 <213> Homo sapien

<400> 84
 ttactacgcg aaatacgagc aggggaacgc cacacagaaa aaaggaaaaa agggagaagg 60
 ggagagagga gaagaaagaa gacggaggga gaggaggag ataggaaggg agggggagaa 120
 cgaaagaaag cgaaaaagg gaggggcaag gagaaggagg gataagagaa aaagaacagg 180
 gggagaagag gaaaagaaga aaaggcagg agagaagaag acagaagagg agaaggcaag 240
 ggagggggac gcacaagaga agaggaggag agggagagaa caaggggaga ggggggggag 300
 ggagaggaga gcagagcggg ccggagaaga aaggagaaag agaaaggga gatgggagcg 360
 acgacaggag gcggggggag ggggagacag ggggaggagg cggaggcggg ggagaagagg 420
 ggggagcagg gggcagtgtg gaggggcaag gagcgagaga gaggggcaag agcgagacgg 480
 cg 482

<210> 85
 <211> 460
 <212> DNA
 <213> Homo sapien

<400> 85
 cttgggctgg gcgcagtggc tcatgcctgt aatcccagca ctttgggagg ctgaggcagg 60
 tggatcacia agtcagcagt tcgagaccag tctggccaac atggtgagac cccgtttcta 120

57

ctaaaactac aagaaagttg gccaggcgtg gtggcacgtg caaattagct attttgggag 180
 actgaggcag gagaatcact tgaaccctgg caggtaaagg ttgcagtgag cctgagatct 240
 gctgccactg cactocagcc tgggtggtgac atgaagtga gattccatct caaaacaaac 300
 aaaaaaaaa aaaaaacaaa aaaaagcgtt tggggtaact ctttggccat acgcgtggtt 360
 ccctttgtgg ggaaattttg ttactccggc tcccacaatc tcccccaac ttatcggggc 420
 aaaatttgat cgtctaaata ctgatctata aatacctga 460

<210> 86

<211> 1161

<212> DNA

<213> Homo sapien

<400> 86

ttaacacact tctaacattt catatataat actataggtc actgtgttat ctagatgcat 60
 actcgagcgg ccgccattaa gtagatagga tcggccgccc gggcagggtc tgcttatcac 120
 aatgagtagt tctacctggt gcagcgtag tagatctttg ccacctatct gtgactttat 180
 gcaatagcat acatgctatt tcatacctaa tagagggagt tccaggagat atcaaccatg 240
 caaatagcat aggatactac aaaggaaaca aacacccaat aaactcggag taggcagact 300
 gacaactgta gagacatagc actatgctac gaaacagaaa tttcatagtt gcaccctatg 360
 tattctacac ctagtagggt tatagacaaa gacaactgcc aaagaatact tcaacgaagg 420
 aggactagca acgtaataat acgtaggtag gagcaacaga agcgacccaa tacaagacc 480
 tagtatctag tacagataga acttgcgata aataactaat agtagctata ctaaggtaag 540
 cgccacatgt ggtctaccac aagagccagt gcctacatta ctacctacat aggctacta 600
 aatccagtac aaatgatatc gtagtaaccg ccatagccct aatacattca taaaacgact 660
 ttctcagcag gcacatcaca actacataat gtatacatta ccaacatact acaacatata 720
 ccataaagtg caatataaaa ctctatgtg tctaataaat aataaaatta gaattaccca 780
 ccttagatgt acaaacacaa cctattcgct aatatctaac cactctctac atacaaaaat 840
 taccaccatc atcaactaaa tcattcaact aaaaatccac aacataacat cctcaaacac 900
 actacacttc cctcacacca tactcaacc acacagtaat attcataaca ctcccaaact 960
 aacaactatc catatctcaa cctccataac tcataccaca ctacaacact atacaacttc 1020
 cttctattct ctactaate gtctccttat actacacaat atatccaa ca taaccacaca 1080
 actcacactt taccacaaac atcaatctct tatacctaca cataacaaca tatcacaat 1140
 tatcacaact aacatactaa a 1161

<210> 87
 <211> 821
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (747)..(747)
 <223> a, c, g or t

<400> 87
 ccgcccgggc aggttgtaat ccagctact tgggaggctg aggcagagaa ttgcttgaac 60
 ccgggaggca gaggttgacg tgagtcgaga tcgtaccact gcactccagc cagggaaca 120
 gaaggagact ccatctcaaa aaaaagaaaa aaaggtaagg ccggactcag tggctcacac 180
 ttgtaatctc agcacttcgg gaggaggctg aggcaggcag attgcttgcg cttaggagtt 240
 caggactgaa ctaggcaaca tggagaaacc atgtctctac aaaatataaa aaaattagct 300
 ggacatggtg tcttgacact gtagtcccag ctactcagga ggctgagctg ggagtatcac 360
 ttgagcccag gaagtgcaga ttgcagtagc caagatcatg cactgcact ccagcctggg 420
 aaacatagtg agatcctgtc tcaaaaataa taataataaa ataggccgag cgcggtggct 480
 cacgcctgta atcccagcac tttgggaggc caaggcgggt ggatcacgag gtcaggagat 540
 caagaccatc ctggctaaca cggtgaaacc ccatctctac taaaaataca aaaaattagc 600
 ccggtgtggt ggtgggcgcc ttagtccca gctactaggg aggcggaggc aggagaatgg 660
 cgtgaacccg ggaggaggag cttgcagtga gccgagattg caccactgca ctccagcctg 720
 tgtaatacag cgagactcca tccaaanaaa aaaaaaaaaa aaaagcgtgg gggaccggg 780
 caacgggtcc gggggaaatg gtcccggcca accaaaaggg g 821

<210> 88
 <211> 716
 <212> DNA
 <213> Homo sapien

<400> 88
 ggagactgca tcatatggcc atgggtccct gatgcatgct cgagcgggcg cagtgtgatg 60
 gatcgccgc ccgggcaggt acggtattgg tgggtgaaat gtaaattagc acaaccacta 120
 tggagaacag ttggaggatc ttcaaaaac taaaaataga gctaccatat gatccagcaa 180
 ttccactgct aggtatatac ccaaaagaaa ggaaattaga tgtggaagag atgtctgcac 240
 tcttatgttt attgcagcac tgttcacaat agccaagatt tggaagcaat gtaagtgtct 300
 accaacagac gaacggataa agaaaagggtg gggccgggcg tgggtggctca tgcttgaat 360

cccagcactt tgggaggccg aggcagatca cctgagggtca gaagtttgag aacagcctgg 420
 ccaatatgga gaaaccccat ctttactaaa atacaaaaat tagctgggcg tgggtggcgca 480
 cacctgtagt cccagctact cgggaggctg aggcaggaga attgcttgaa cctgggaggc 540
 agagattgca gtgagccaag attgtgggca acagagcaag gctccctctc aaaaaggagt 600
 aaataaaaaa aaaaaaaaaa aaaaaaaaaa aaaggctggg ggtaccgggg ccaaaagcgg 660
 gttcccgggg ggaaattggt tttccgcca aaattcccc atatgcaaaa aaggga 716

<210> 89
 <211> 523
 <212> DNA
 <213> Homo sapien

<400> 89
 gcccgggcag gtaccataaa tcacaggctg agggagaaat ggtgagggca caatagcaaa 60
 tggaaatata caaaaaatag ctgggtgcgg tggctcacac ctgtagtccc agcactttgg 120
 gaggccaaga taggcagatc acttgaggcc aggagtcca gaccagcctg gccaacatgg 180
 caaaacctg tctctaccaa aactgcaaaa attagctggg tgtggtggcg tgcacctgta 240
 tcccaactac tcgggaggct gaggcataag aattgcttaa acctagatgg cagacactgc 300
 agtgagctga gatcatgaca ccgcactcca gccctatgt aacagagcag actctgttcc 360
 gaaaaaaaa aaagaaaaaa aagtctgggc ggtagatctt gggtcctaaa gctggttccc 420
 tgggtggtgaa tatttggttt cccgctccac atattccaca caacaacgga accaagggtc 480
 tgttcacata ccattgttct ggtggagacg tcagctgaca cca 523

<210> 90
 <211> 673
 <212> DNA
 <213> Homo sapien

<400> 90
 tgggtgtcagc tgacgtctcc accagaacaa tggatatgta acagaccctt ggttccgttg 60
 ttgtgtggaa tatgtggagc gggaaaacca atattcacca ccagggaacc agcttttagga 120
 cccaagatct accgcccaga cttttttttt tttttttttt ttcggaacag agtctgctct 180
 gttacatagg ggctggagtg cgggtgcatg atctcagctc actgcagtgt ctgccatcta 240
 gggttaagca attcttatgc ctcagcctcc cgagtagttg ggatacaggc gcacgccacc 300
 acaccagct aatttttgca gttttggtag agacagggtt ttgcatgtt ggccagggtg 360
 gtctcgaact cctggcctca agtgatctgc ctatcttggc ctcccaaagt gctgggacta 420

60

cagggtgtgag ccaccgcacc cagctatattt ttgtgtattt ccatttgcta ttgtgccctc 480
 accattttctc cctcagcctg tgatttatgg tacaatatcc tgaagatgtc gggagcacac 540
 gacctcagcc tcggctgcag atagactgct cagcttgggg caataactgc tctgccctcc 600
 ctctgccacc cggcagcccc acccacagaa gggcccagac ttacggcttt ggagggagca 660
 tagtgtgtcg gtg 673

<210> 91
 <211> 744
 <212> DNA
 <213> Homo sapien

<400> 91
 aagaggatgat gactcactat ggccctgtta ctctagatca tgctcgagcg gcgctagtgt 60
 gatggattgc cagggccata tcctcctacc acaggcgaag ctggatagca gaggagatgg 120
 ggagatggga gaaggacggc tgactgagtc acggttggcc tgggtggctg cagaaaagaa 180
 aaaaaaaaaa aaaaaaagaa aaaaaaaaaa aatgggggaa aaagggcaca gggagtccg 240
 gggggaaatt ttcccgggcc gcattctccc ccaaaaatat ataaggaaat aaggtcaata 300
 agaaagatta tagaaaaact agaagatata gggtgaaaga gtgatatgat caagaaaatg 360
 aatcaagagg aacgatagtgt ataagtagaa atcaaagaac aattgaaaga taaaggaaac 420
 aaaaaaatga agagatgaga ataaaaaatg gaagataaac attgaatcaa tgaagactaa 480
 aaggagaaat cgaccatcca cactatgaga gaagacatac ccacaataca agagagaaga 540
 acaatagaga gagacagaaa acgaacgacg aaaggcataa aagaacgaaa agagtaaaga 600
 gaaaatcaaa cgcaagagcc agacaaaaca aaaacaagaa acacagcaag aagcaaaaaa 660
 aacgacaaga gataaagaaa gaaaacaaga aggtagaaaa agagaaagaa aaaaaaaaaa 720
 aagataagaa taggaaagac aaac 744

<210> 92
 <211> 879
 <212> DNA
 <213> Homo sapien

<400> 92
 agtcccttcc cgtgatgtga aatgcagtgt gggctggagc ccttgtatgc tttgccctca 60
 gtaggcgctt gcctcccctc actgggcact gaggccatag gccgctttgt tctgcagacc 120
 acgcctggcc tctagaggaa tggatgtgac ctagagatct ggactcagag ctgggcagaa 180
 cctgggtgat gcttactgag caggatgccc aggtttgtga ctgtgtctat gagaggcctt 240
 aaggcatgta ggggtcatct gggggaaagg acggctgact gagtcaagggt tggcctgggt 300

ggctgcagaa aagaaaaaaa aaaaaaaaaa aagaaaaaaa aaaaaaatgg gggaaaaagg 360
gcacagggag ttccgggggg aaattttccc gggccgcatt ctcccccaa aatatataag 420
gaaataaggt caataagaaa gattatagaa aaactagaag atataggggtg aaagagtgat 480
atgatcaaga aaatgaatca agaggaacga tagtgataag tagaaatcaa agaacaattg 540
aaagataaag gaaacaaaaa aatgaagaga tgagaataaa aaatggaaga taaacattga 600
atcaatgaag actaaaagga gaaatcgacc atccacacta tgagagaaga catacccaca 660
atacaagaga gaagaacaat agagagagac agaaaacgaa cgacgaaagg cataaaagaa 720
cgaaaagagt aaagagaaaa tcaaacgcaa gagccagaca aaacaaaaac aagaaacaca 780
gcaagaagca aaaaaaacga caagagataa agaaaagaaa caagaaggta gaaaaagaga 840
aagaaaaaaa aaaacaagat aagaatagga aagacaaac 879

<210> 93
<211> 676
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (489)..(489)
<223> a, c, g or t

<400> 93
gcggccgccc gggcaggtac ccaacagctc attgagaacg ggccaggatg acaatggcgg 60
cttgtggaat agaaaggcgg gaaaggtggg gaaaagatga gaaatcggat ggttgccgtg 120
tctgtgtgga aagaagtaga catgggagac ttttcattct gttctacact aagaaaaatt 180
cctctgcctt gggatcctgt tgatctgtga ccttaccccc aaccctgtgc tctctgaaac 240
atgtgcggtg tccactcagg gttaaattga ttaagggcag cgcaagatgt gctttgttaa 300
acagatgctt gaaggcagca tgctcgtaa gagtcacac caatccctaa tctcaagtaa 360
tcaggacac aaacactgcg gaaggccgca gggctctctg cctaggaaaa ccagagacct 420
ttgttcactt gtttatctgc tgacctccc tccactattg tccatgacc ctgccaaata 480
cccctctgnt gagaaacacc caagaattat ctaaaaaaaaa agaaaaaaaa aaaaaaaaaag 540
gcggggggaa accagggcca aaggggttcc gggggcgaaa ggggttctcc gcaccaaatt 600
tccacaaaaa taggagcaaa gaaaaagaaa gaaaaaaaaa aaaacaaaaa aagaaaaaag 660
aaaacaagag aaagaa 676

<210> 94
 <211> 850
 <212> DNA
 <213> Homo sapien

<400> 94
 cgcccgggcg ggtacgtgct ccgggatctt gagccaacca tggccggcat cggcgtgact 60
 tggttttctca cgtgtgactc atctgcgcca acttgtgccc gatatacacag ttccggctcga 120
 ccatcgatgc ctgaagctcg acgaaggta tggataatac agatacaagg tgatcccaag 180
 aaaatgggtc ctctccata ggaggtcgaa gaaggatagt gaagatcagg aacattgtcg 240
 cggaaagcat gatacgaact atgataagta gtggagaaga ggtctgtcag tacctcatga 300
 gatgcaatag gctaggaacg gcgggtgcaa actctgcagt acaggatagg tggcccgcta 360
 tctccccaat cacaagcagt tgcagttgcc acaccagcca agaaaaaaaa aaagaaaaaa 420
 aaatgggctg gggggggata cactacgtgg gggccaatag gcgagctacc cccggtggtg 480
 tgagaatgtg ggggtgttgc gcggccacca caatttgcgc gccaccaagc attagcggag 540
 ccgaaacagg gcagaagggc aggggaagcaa cgtaggcgta gagcaccctg gggggcaggg 600
 gcaacgaagg acttggaag gagacagacc caacggaact cggatgaggg ccaccgcaag 660
 acatgacgaa agaaacattc ccacaccgtc gacaacacag ggagcctaaa aactacaga 720
 aacacaacgc aacaagcaaa aacaactaca aaaccaccag cggacctgag agcatacact 780
 aagtggatca aacaagaggg aaatcaccag aggaaaacaa aaaaaggaca aacagcacac 840
 aaacacacac 850

<210> 95
 <211> 644
 <212> DNA
 <213> Homo sapien

<400> 95
 cgagcggccg cccgggcagg tacaagcgtt tttttttttt tttttttttt ttgggaaagg 60
 gatttttctc cctgttgccc aggttgaggt gcaaacatga tctcggtctc acggccccct 120
 ccgccgttct tctgcgggtt caagcaattg tttggcctca gccccgcga ttagtcctgg 180
 agattacagg gtgcggcaca taccacacca aggctaattt tggggatatt ttaaggtaag 240
 agatgggggt tttcaccatc tgggccaggg ggggtcttga atccctgacc tcgggtgatc 300
 caccacctg gtgcctccca aagtgtagt attatggggc ggtagaacca ccatgccac 360
 gccgaaaacg ctgggcggtg actcatgtgg ctcaaaaagc agtggttccc cgtgggtggtg 420
 aaaaagttag gttatctccc gccctcatca cattctccac caacaaccaa taaccagaaa 480

63

gacaaaccgc gggggggggg gggcaaaggg cgggggctga ggcacacaaa cagaggaagg 540
ggaaagacaa aagacaaaga ggaaaaagag gcgagggact aaagaggggg gcgaaaaaaa 600
acaaaaaaa aaaaaaaa aaaaaaatag cagaagaaga aagt 644

<210> 96
<211> 846
<212> DNA
<213> Homo sapien

<400> 96
ggcgcgatt tttttttttt ttttttgtga ttttaaaaa cagaaaactt tatttgaaca 60
agaaaaagtt aaaaatgtta cacttcgaaa aaatttttaa ctcgttaatc atttttaatt 120
gacaaataac tgcatttaca gggacttaca tgactttttg atacatgtga ttaaaccagg 180
ctaattaaca tatccatcac ttcacttttt ttgtggtatg aatattttaa atctctctta 240
gcaatttcct tttttttttt ttttttggg acagagtctt gctctgtcac ttagactgga 300
gtgcattggt gctgtctcca ctactgcaa cctccgcctc tgggattcaa gcaattctcc 360
tgcctcagcc tcccaaatac ctgggactac aggcattcac taccatgcc agataatttt 420
tgtattttta gtagagacag gggtttccga gacagggttt caccatgttg gccaggctga 480
tcttgaactc ctgacctcag gtgatccacc caccttggcc tcccaaagtg ctagtattat 540
gggcgtgaac caccatgcc acgccgaaa cgctgggcgg taactcatgt ggctcaaaaa 600
gcagtgggtc cccgtgggtg tgaaaaagtg aggttatctc ccgcctcat cacattctcc 660
accaacaacc aataaccaga aagacaaacc gcgggggggg gggggcaaag ggcggggggt 720
gaggcacaca aacagaggaa ggggaaagac aaaagacaaa gaggaaaaag aggcgaggga 780
ctaaagaggg gggcgaaaaa aaacaaaaaa aaaaaaaaaa aaaaaaaat agcagaagaa 840
gaaagt 846

<210> 97
<211> 1604
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (202)..(202)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (562)..(562)
<223> a, c, g or t

<220>
 <221> misc_feature
 <222> (1313)..(1313)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (1570)..(1570)
 <223> a, c, g or t

<400> 97
 cgagcgggcg gccgggcggg tacagtggga cttgctggca ttcgaggccc tcggggttca 60
 ccaecgggccc tgtgggtccc ccttggtccc cctgggccct cctgggagct ccaggctagt 120
 aagacgcgct gggttggatt attgagtatt gtgttaactg agtaggatgg agctttctag 180
 cagagggcctt gagcccaggg gnttcggctc tacacaaacc tctctgctca atcaccgcta 240
 gaggcacgta ctgagacgtc tggctcgtc tatcatcatg atagcgtcgc tcatctaaca 300
 agccaggaga ttgacagacc catctctgta gctctcatga taggagcatc ttatgaaata 360
 gaaatccgca gtcttcgtgc cctagtgcgc cgtgagcttt aggagaatca tacctggcac 420
 aatcctccag gagataggta aggcaggtta ggctatgac gtgatcgtgg gtaatctgga 480
 tccgctaata acacgaatgg gaagtgtccc ctggatagag agtggcatag tcctaaagtc 540
 attacgttgg tgcattgcgc cntccttata cctcggggcc gtacaaacgc tgtgggctta 600
 tacctttgtg agcacgcgca tacccttgag gagacaaaca tatcgacct catggcgcta 660
 aagcggaaac tttggtgtta taatgggaag cttcccgaaa ttgggaacca aagaaaaaat 720
 actaatctta tgtgtcttga cggtgagggc atgtaaacgt tattaccaca tttaagatcg 780
 tgtgggataa cggtgggcca atgtggatgt ggatattaat atattaggtc cgtctatgaa 840
 ataacgggga ctgtttgaat cctttttccc aaagggggta aaaaactggg gtgccttatt 900
 ccgctaaact tcttttgtgg gcgcccttat accatatttg gtcctgctcc ctagtggctc 960
 tgtggagccc cgatgggcat ttatttgctt cccattctc ttccaaccgt aagaggctaa 1020
 actccactct cggttacacc ccacccttg cgcaaagtgg attatcccat tgccattgtg 1080
 tgtgtccctt acccagtggg cgaccctgc aaccgggggt gtacgcctct ggttggcggc 1140
 gagcccccg ttggtggagt acacagcggg gcgtggcccc aattgggcct tcaccagcgg 1200
 ggcgcctgtc cttaaagtag gaatttccct ttggaaaccc catgtgggggt tgaggcactt 1260
 ggagaagacg gggcggaac cacgaccggg ggagtttcca caccttgtgt tanccagcgc 1320

65

gtgggggtccc agttttgggt gagaggaaaag agggcggggt ggggcttcgc taaaaaggag 1380
 acggaacgga tcgttgcgca gcctggcgtc gggcgacaaa ggcgactacg gttcacattt 1440
 tgcgtcatct tccacccggc ccgcgcacaa ttcatagccc gaggtcacag cgccgtgttg 1500
 gccccgattc ccttgtgagt tattgtggcg ccccttttgg ggaccacact cctggggggc 1560
 ggcggcgtn acaccaggag aacatccctt ttgcgtggca cacg 1604

<210> 98
 <211> 2158
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (756)..(756)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (1116)..(1116)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (1867)..(1867)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (2124)..(2124)
 <223> a, c, g or t

<400> 98
 tgaacgtggt ctaccagggt ttgctgggtgc tgtgggtgaa cctggtcctc ttggcattgc 60
 oggcccctct ggggcccgtg gtcctcctgg tgctgtgggt agtcctggag tcaacgggtgc 120
 tcctgggtgaa gctggtcgtg atggcaaccc tgggaacgat ggtccccag gtcgcgatgg 180
 tcaaccgga cacaaggag agcgcggtta ccctggcaat attgggtccc ttggtgctgc 240
 aggtgcacct ggtcctcatg gcccgtggg tcctgctggc aaacatggaa accgtgggtga 300
 aactggctct tctggctctg ttggctcctgc tgggtctgtt ggccaagag gtcctagtgg 360
 ccacaaggc attcgtggcg ataagggaga gcccggtgaa aaggggcca gaggtcttcc 420
 tggcttaaag ggacacaat gattgcaagg tctgcctggt atcgtggtc accatggtga 480
 tcaagggtgt cctggctccg tgggtcctgc tggctcctagg ggccctgctg gtccttctgg 540

66

```

ccctgctgga aaagatggtc gcactggaca tcctggtaca gttggacctg ctggcattcg      600
aggccctcag ggtcaccaag gccctgctgg ccccccctggt ccccctgggc cctcctgggg      660
acctccaggc tagtaagacg cgctgggttg gattattgag tattgtgtta actgagtagg      720
atggagcttt ctagcagagg gcttgagccc aggcgnttcg gctctacaca aacctctctg      780
ctcaatcacc gctagaggca cgtactgaga cgtctggctc gctctatcat catgatagcg      840
tcgctcatct aacaagccag gagattgaca gacccatctc tgtagctctc atgataggag      900
catcttatga aatagaaatc cgcagtcttc gtgccctagt gccgcgtgag ctttaggaga      960
atcatacctg gcacaatcct ccaggagata ggtaaggcag gttaggctat gatcgtgatc     1020
gtgggtaatc tggatccgct aataacacga atgggaagtg tcccctggat agagagtggc     1080
atagtcctaa agtcattacg ttggtgcatt gcgcctcctc tatacctcgg ggccgtacaa     1140
acgctgtggg cttatacctt tgtgagcacg cgcatacccc tgaggagaca aacatatcgc     1200
acctcatggc gctaaagcgg aaactttggg gttataatgg gaagcttccc gaaattggga     1260
accaaagaaa aaataactaat cttatgtgtc ttgacggtgg aggcattgaa acgttattac     1320
cacatttaag atcgtgtggg ataacggtgg cccaatgtgg atgtggatat taatatatta     1380
ggtcctgcta tgaaataacg gggactgttt gaatcctttt tcccaaaggg ggtaaaaaac     1440
tggtgtgcct tattccgcta aacttccttt gtgggcgccc ttataccata tttggtcctg     1500
ctccctagtg gctctgtgga gcccgatgg gcatttattt gcgtcccat tctcttccaa     1560
ccgtaagagg ctaaaactcca ctctcggtta caccacccc cttgcgcaa gtggattatc     1620
ccattgccat tgtgtgtgtc cettaccagc tgggcgaccc tcgcaaccgg ggggtgtacg     1680
ctctggttgg cggcgagccc cccgttggtg gactacacag cggggcgtgg cccaattgg     1740
gccttcacca gcggggcgcc tgtccttaaa gtaggaattt ccctttggaa accccatgtg     1800
gggttgaggc acttgagaaa gacggggcgg aaaccacgac cgggggagtt tccacacctt     1860
gtgttancca gcgcgtgggg tcccagtttt ggtgagagg aaagaggcg ggggtggggct     1920
tcgctaaaaa ggagacggaa cggatcgttg cgcagcctgg cgtcgggcgc acaaggcgac     1980
tacggttcac attttgctc atcttcacc cggtccgcgc acaattcata gcccgaggtc     2040
acagcgccgt gttggccccg attcccttgt gagttattgt ggcgcccctt ttggggacca     2100
cactcctggg gggcggcggc gtgnacacca ggagaacatc ccttttgctg ggcacacg     2158

```

<210> 99

<211> 1034

<212> DNA

<213> Homo sapien

<400> 99
 tcggtgccc gaggtctcgt gagcccccta ggaccatcac ggatgccgag cttcggggta 60
 actctcagca gtgtgcaagg ttcccaactgc ctgccgccgt aaatccctgc tcgaaagtca 120
 cgctcctgga cgaaacagtc ggccgcactt catctgcgtc cagtatcaca ctccccataa 180
 tgagttgtct gccacctcca agcacttcaa cactatttcg ttttattttt ctgattagtt 240
 ataacgacgg caggaatgtc taggccgtct gagcccaggc caagccatct gcacccctct 300
 ttgacattgc acggtatatg cccagatggc ctgaaggtaa cttgaagaat caccgaaagg 360
 aagtgaatat gctctgcccc tacctttaac atgtatgaca cttcctacct acaagaagag 420
 agtgtcaaat ggccggctct tgctttaagt gatgacatta cctgtgggtga aagtccttat 480
 tcgctgtgct catcctggct caaaaatcac cccactgag caccttgcaa cccccactcc 540
 tgccctgccag agagcaaacc ctctttgact gtaattttcc tttacctacc caaattccta 600
 taaaacggcc ccacccttat ctcccttcgc tgactctctt ttccggactca gcccgccctgc 660
 acaccagcga tgaacataaa caggccacga ttgcatcaca acaacacaca cacaacacac 720
 acaaacaccc ggagagacaac cagtgcacca caccggggcc cgggggccac caggtaaccc 780
 ggccaaatcc ccaaatacta ccactcaca agtccacata tcaagcatct caacacacac 840
 ctcttaccac ataccataaa catacacaca cccactcacc caccacctca cccactcaca 900
 acacaccta caccactctc acccccccca cccacatctc aacaccacac ccaccatcca 960
 acaaccaccc catcacacca cccccacacc ccaccctcac ttaacacaac aactctcccc 1020
 accatcccc ctcc 1034

<210> 100
 <211> 1401
 <212> DNA
 <213> Homo sapien

<400> 100
 tcggtgccc gaggtctcgt gagcccccta ggaccatcac ggatgccgag cttcggggta 60
 actctcagca gtgtgcaagg ttcccaactgc ctgccgccgt aaatccctgc tcgaaagtca 120
 cgctcctgga cgaaacagtc ggccgcactt catctgcgtc cagtatcaca ctccccataa 180
 tgagttgtct gccacctcca agcacttcaa cactatttcg ttttattttt ctgattagtt 240
 ataacgacgg caggaatgtc taggccgtct gagcccaggc caagccatct gcacccctct 300
 ttgacattgc acggtatatg cccagatggc ctgaaggtaa cttgaagaat caccgaaagg 360
 aagtgaatat gctctgcccc tacctttaac atgtatgaca cttcctacct acaagaagag 420

68

agtgtcaaat ggccggtcct tgctttaagt gatgacatta cctgtggtga aagtccttat 480
 tcgctgtgct catcctgggt caaaaatcac cccactgag caccttgcaa cccccactcc 540
 tgcttgccag agagcaaacc ctctttgact gtaattttcc ttacctacc caaattccta 600
 taaaacggcc ccacccttat ctcccttcgc tgactctctt ttccgactca gcccgctgc 660
 acaccagcga tgaacataaa cagccttggt gctcacacaa agcctgtttg gtggtctctt 720
 cacacaaacg cgcagaaat ttggtgccat gactcgatc ggggtacctc ccttgggaga 780
 tcaatcccca gtccctcctgc tctttgctcc gtgagaaaga tctacctagg acctcaggtc 840
 ctcagactga ccagcccaag gaacatctca ccaatttcaa atctggaacg cgcagaaaa 900
 aaccaaaaaa acaaaaaaat tcttttggtg gcagaataaa aaaacaaaaa aaaggacttt 960
 ttcttctgga ctgaactata tttaaactct aaaggatgga catctcacia ccttctaca 1020
 gcaagtactg tgagagctgc atcttgtccc actggatggt cttcagagac aataatacat 1080
 aatggagctg tcatctccta tgataacaat gccttcttct ggatacctcc tgaaggacct 1140
 gcctgaggct ctttccactcc atgaaaagggt gcgctgtttc ccttgtgcct tccgcattat 1200
 gaaagtttcc tgaggcctcc ccagcatgct acctgtacga ctgtgaaaca taaggcaaaa 1260
 aaagatacta gcgccagttt ggaggggctt ttctagtctc aagcacgctg aggagtctaa 1320
 cttggctttg gagagaaagg gaagctagct ctgagctgga ggggcttagg ggggtctctg 1380
 gatccgaggt taagggcgga g 1401

<210> 101

<211> 952

<212> DNA

<213> Homo sapien

<400> 101

gcggccgccc gggcaggtag ccaacagctc attgagaacg ggccaggatg acaatggcgg 60
 ctttgtgga tagaaaggcg ggaaaggtag ggaaaagatt gagaaatcgg atggttgccg 120
 tgtctgtgtg gaaagaagta gacatgggag acttttcatt ttgttctaca ctaagaaaaa 180
 ttctctgcc ttgggacccg gttgatctgt gaccttacc ccaaccctgt gctctctgaa 240
 acatgtgcgg tgtccactca gggttaaatg gattaagggc agtgcaagat gtgctttgtt 300
 aaacagatgc ttgaaggcag catgctcgtt aagagtcac accaatccct aatctcaagt 360
 aatcagggac acaaacactg cggaaaggcg cagggtcctc tgcctaggaa aaccagagac 420
 ctttgttcac ttgtttatct gctgacctc cctccactat tgtcccatga ccctgccaaa 480
 taccctctg tgagaaacac ccaagaatta tctaaaaaaa aaaaccacia accaaaaaaa 540

69

```

aaaaggctgg gggacccatg ggccatagct tgtccctgtg gtggcattgg taccgcgcaa 600
ttccccattg tacaacaaac actccaacta ctaccacact gctaccaaca taaacaaata 660
gactcctctc gcatctatcc cctgcaaata aattaattca actataatgc acacaaaaac 720
aaacccttta agtaaatac acctctactc acaataaaaac tgtcacaacc tatcatccta 780
tcactacaca ccatcaaac caagtcaaca ccatgaacac caacacaaaa tacacaaaaa 840
acatacacta acacactaac atattcatac tacaccataa ctacacacag accaacaaaa 900
aacaacact acgcactaca ctcacatata ctacttacia ctccactcaa ct 952

```

```

<210> 102
<211> 1549
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (844)..(844)
<223> a, c, g or t

```

```

<220>
<221> misc_feature
<222> (1217)..(1217)
<223> a, c, g or t

```

```

<400> 102
gcgtggctgc gcccgagcgt gccagcgcag gtgcttctgc tgagtggggc aggcggagct 60
tgacgaaacc gcctataacg ttctttttct ctcttgacca cgagtagagc acttaagtac 120
cagctactta aagaaagtat agtcaatag agtcactaca gaatattagc ttaggcgtta 180
gacgttttta gacgctaagt tctgtaacta cgctgtaacg catttttcaa cgcagaagaa 240
taatgacatg actgtagaag cgacgtagca gtggacggac aggaacacac gatacacata 300
gggtcttcta gaaaccatgt gacctggacg cgttgtgcag gattgaacgc cttgcgccgt 360
gaatccctgc gtcacttagg ctaggttttc catgtgatcg acaccgttgg tttaactccc 420
cagacgtcga ccgataatct cctgacgagc gagagcatac ccgtgaagtc caaagcctag 480
aacgaatagc acgactaccg tgggacagtc gaagtggcaa aggccagaga tatcatttgg 540
gtttgatcca gcaacgagtg ccgtatttcc acaccgtacg agtatgctcc caaagagtat 600
atagaggggc gccccaaaat acggtcacac gtaatatgtg ccactcccgt gtgctagagc 660
aaaacgatac acacgctgct ccatttgtga acaccttctg cgccgcaata aatgggtgat 720
acataagagc acatttctaa ccacagagta gtgcagtgca atggcccat tctcagaagc 780

```

70

atcccccata	agagctcaat	acattccttg	ttcctgtgtg	acagagcttt	aaaacaattt	840
gacncttctc	ctttaatcta	agagggtag	gcagcagtgt	agttgcgaag	cctaactgct	900
caagagatag	ttgaatcaaa	taccgcccc	cggtgtcttg	gggccacata	ctggagaacc	960
atttcccgcg	gtgtcaaata	atatggagga	accgtaaact	caaacttgtg	aggctttcgc	1020
cttaaagccg	cccggatact	aaaacggacc	gaacacaacg	catggcctta	aagtcgttgg	1080
ttaacataga	aacacacaag	aacccccatg	gtagccac	gatatgtttt	tgagcaccat	1140
accgtgatga	tatcgtaagg	atttgcacgc	cattgttgtt	gtaaagacat	gtgatgtcgg	1200
tggatatgta	aacaacncac	atttacccta	agggcaccaa	aggagcgtaa	tgaagacaat	1260
gaggccaccg	actttcccc	catagggcag	ccaacggcac	ttgtgtggca	cactcgcgtt	1320
taagcgact	ctaccatagt	atgaggacgc	ccgctcgtct	ttgagtagaa	gagacctaa	1380
aggagagcaa	caacaaaagc	taaacagaaa	ctaggagcag	caaaacaaaa	ccaaagcaat	1440
agcgacagcg	agagacagaa	gtgtactcgt	gcaccactaa	ggacacatgt	cagtgtggg	1500
ggtagtccag	acagatgtga	ttcaccctgt	acggcagccg	cacagacag		1549

<210> 103

<211> 767

<212> DNA

<213> Homo sapien

<400> 103

atgaattcat	ctatagggcc	attgtgactc	tagatgctgc	tggagcggc	gccttgtgat	60
ggatcgccg	cccggcggt	tgtaatcca	gctacttggg	aggctgaggc	agagaattgc	120
ttgaaccgg	gaggcagagg	ttgcagttag	tgcagatcgt	accactgcac	tccagccagg	180
gcaacagaag	gagactccat	ctcaaaaaa	agaaaaaag	gtaaggccgg	actcagtggc	240
tcacacttgt	aatctcagca	cttcgggagg	aggctgaggc	aggcagatgc	ttgcgcttag	300
gagttcagga	ctgaactagg	caacatggag	aaaccatgtc	tctacaaaat	ataaaaaaat	360
tagctggaca	tggtgtcttg	cacctgtagt	cccagctact	caggaggctg	agctgggagt	420
atcacttgag	cccaggaagt	gcagattgca	gtagccaaga	tcatgccact	gcactccagc	480
ctgggaaaca	tagtgagatc	ctgtctcaaa	aataataata	ataaaatagg	ccgagcgagg	540
tgactcacgc	ctgtaatccc	agcacttttg	gaggccaagg	cgggtggatc	acgaggtcag	600
gagatcaaga	ccatcctggc	taacacggtg	aaaccccatc	tctactaaaa	atacaaaaaa	660
ttagcccggt	gtgggtgggg	gcccctgtag	tcccagctac	tagggaggcg	gaggcggaga	720
atggcgtaaa	cccgggaggg	tggagcttgc	agtgagccga	gattgcc		767

<210> 104
 <211> 635
 <212> DNA
 <213> Homo sapien

<400> 104
 cgagctggcc gccccgggca ggtcacacct taccacttga caccattaa caaaactacc 60
 cctgtccctt attccctgtt gcccctcttc gtgaacgggtg gatcatagt cgtcgttttc 120
 caaccgtttg ccaaactctt tgcacatgtc cctcctctct gcgcatgaca tcaggagctt 180
 gtgagcgtgg aagtacctaa cctagctcat gactaccaga acgttccttg tatagaaaga 240
 ctctacacct attctgagtt ttcaaagtat gactgatccc ctggggcagc gtcgaaaggc 300
 gtttggccgc ttaaaactcca atcgcgctca tcaggcttgg ttccccctag tagttgcaac 360
 attccgtttc actcccgctt caccatagt tccccagcga cgaatccatc acttgagggc 420
 cactccaact aggaggttta aggtcgaccc caggggggat ccttggcacg tgaaccctt 480
 ttgacacagt tttttgcaca ttgaaacttt gacagctttg agtaatttct cottacgaga 540
 acccccctgg cgagctttct tataccatcg ttttgcggag ctccattata tataacggta 600
 caacacaggt ttttttgtca accccgtcct gaatt 635

<210> 105
 <211> 461
 <212> DNA
 <213> Homo sapien

<400> 105
 cacaacatac gagcatgact tggagaggtc tcatttgctg aatgagtgat gaatgctgac 60
 cacacattca aggaaaacct tggatgttca tacatgtatt ttaaaactga gggatatgtc 120
 ctttctgaga gatgtatcaa gcatggataa ctgaaagggt tttgagtgtc taaaacggat 180
 aagctccaga atatctggaa ccattcacat tgcataatgt cactacatta tcccagagta 240
 gtagtttgtt aaagttaaca cgttagtgtt aagctaaagt gctagagggt cgtttttcgc 300
 tgttctagac gagaggtgaa tagtcataaa gtcaagttca ttagaacgag gaaaaaacia 360
 aaaaaaacia gaaacaaaaa aaaaagggtg gggtaaacia atgggaaaaa aggggaccct 420
 ggttgaaaaa ttagttaccc cagacaaaaa attcccagca a 461

<210> 106
 <211> 1041
 <212> DNA
 <213> Homo sapien

<400> 106

72

tcgcggccga ggtaccacca ttgaaaacat ttaagttggc caggcacagt ggtgcacgcc 60
 tgtaatccta gcactttggg aggccgagggc aggtgagctg cttgagccaa aaagttccag 120
 accagcatgg gcaacatgca gaaacccgt ctctagaaaa tatacaaaaa ttagcggggc 180
 atggtggcac atgcctgcag tctcagctac tcaggacaat gaggtaggag gattgcttga 240
 gcctgtgaag ttgagactgc agtaagctgt gatgatgcca ccgctctcca ccttgggtga 300
 cagagcaaga cccgagaaag aaagaaagaa agaaagaaag aaagaaagaa agaaagaaag 360
 aaagaaagaa agagagagag agagagagag agagaaagaa agaaaggagg aaggaaggaa 420
 ggacggacac ggacacggaa cggaacggta cggtaacggc acacggaccg gggaaagcaa 480
 gaaagaaaga acagaaaaag aacgacagac cgacagaaag aagagagaga gaaagacaga 540
 acggaggtcc ggacggaccg gacggacgga caaaaagaag aagaagacgg aacgacagaa 600
 cagaccgacg cgacagaagc acgaagaccg acagacgagg gaccgaccga ccgaccgcac 660
 cgacttccat aaataaaaag gcgtgcgagg aacaaggtgc caataggggtg accgtggggg 720
 aaatgtgtct cggccaaatc aacggaggtc cacacaaagc tgatggaagt caaaaaaaaa 780
 agaaaaaaga ggatagaaga aaaagagtga gacaagaaaa agaaggggaa aaccagggat 840
 ggaaggggaa ggagagacag aggagaagag caaggagggg agaccaagga ggaagaagga 900
 gaggcggaaa cagggggggg aaaggagaag aaagaagcaa agagaacagg gaaaggaaaag 960
 aaggaagacg aggacacaga agagaaggcg aaaagacaga gggacacaag aacaagaaga 1020
 gcgcgagagg aaggaggagg g 1041

<210> 107

<211> 834

<212> DNA

<213> Homo sapien

<400> 107

tggtcgcggc cgaggtacac gcccgcctcc gggagagctg agtcaggccc tgaaaagcct 60
 caccttgaac catgtggtct aggcaagccc ccgagcctcc tctggccgat aagttgacct 120
 ccccaaactt ggtatggtag ccctggggct tgtctgcttg tgagtatcgg ggagaaccgg 180
 cgtggttgat ggggggcccc cagcagctgc taatcagttt agacgactct atatcccgct 240
 gcagaatcaa ctgcctactg agaggcaagg ccagccttcc accggaagga gaagagccag 300
 tataatgggt ggccagtgc gcgttcgtgc ttggtgcca ttcttgagtc aggggtgtcat 360
 ccgttgatgc attcatgacg cactgctgga gagagaggct gcatgagctt cccctagaac 420
 agttgaaact aaaagacttg tgtgccgttt aaaaaataac acaataatac catataataa 480

73

ggcgtggggg gtcaaccag tggcagccat ggcgtgatt cccggtgggg ggtagaatgc 540
gtgtcgccgg ttcacaatct tccagataga cttttagaga gcacaagggt taaaaatggc 600
cctttcaaaa attaaacctg tcaaaaacac aaaagagaaa gaacaaacaa aaagaacaaa 660
gaagagggag tggggggaaa caccgggggg cacaaagagt gaaacccggg tgagacacac 720
tgggataacc ggcaccacaa ttcccaccaa actaaagggc gcaaagagag aaaaagaaga 780
gcgaaaacaa aaaaaaaaaa acgaagagga gaaaaaaaag agaaaaggaa gaag 834

<210> 108
<211> 1015
<212> DNA
<213> Homo sapien

<400> 108
agttcgacaa tataggggcca tggttatcta atgcatgctc cgagcggcgc cttatgtgat 60
ggatcgggccg ccgggcaggt gcaaaaatca gcccaacatg gtggtgcaca cctataatcc 120
cagctactcg ggaggttgag gcacaggaat cgcttgaact cggtaggcga aggtttcagt 180
gagccaaaat catgccactg cctccagcc tgggcaacag agtgagaatg tctcaaaaaa 240
aaaaaaaaaa aaaagagaaa aatggggggtt ggaactacac agggtcctcc ttataaggca 300
ggattctttt caattaaaag ttacaccaag gtgtgcctgc ctctccttcc cggttttctc 360
cacctcttcc atcccttgct tacctcgggg ggggaaaga ccaaaccctc ctcttcatct 420
ctcctccaca gcctactcag tgcaaagacc gtgagggatg aagactttag tgatgatcca 480
ctttccactt aatggaatat gtaaaatatt tggctcctg gtatagaatt tccctaaata 540
accattgggt tcttggtagc tatactttac ctggcaagaa tatacggtt ggtaatactt 600
attaaccata ccggaatatg tggcactgg tatggggtaa gggcttctgg tcagcaatag 660
ggctattagt aagttggggg ggagtcaaca cggatatggc ggggtttccg acttaactta 720
atatattcct tttatcagcc acttatttaa tgtcccacca gcatattcct tgactcactt 780
tttattgggt gattgggggc ctacaggtat gttccctttt aagattcagg ttgcacaggt 840
aaggcgatat ggtgagccgg aaggtatctg ggacaagaga ctgtgtggaa atgacccggg 900
agagattcct ttacaatgat tcaggtgggg gaggagccaa tgtgcatatg gaagcgccgt 960
aagccccagg agcacacgaa agggttcgcg ttccactaaa acggttcccc gcgtg 1015

<210> 109
<211> 577
<212> DNA
<213> Homo sapien

<400> 109
gcgccgggca ggttgtaatc ccagctactt gggaggctga gtgcagagaa ttgcttgaa 60
ccctggtagag gcagagggtg cagtgatgtc gagatctgta ccaactgcact ccagccagtg 120
gcaacatgaa ggagactcca tctcaaaaaa atgaaaaaaa ggtaatgtgc cggactcagt 180
tggtctcacac ttgtaatctc agcacttacg ggaggacggc tgagtgcagg cagattgctt 240
gcgcttagga gttcacggac ttgaacctat tgcaaccatg ggagaaacca tgtctctaca 300
aaatatataa aaaaattatg ctggacatgg tgtcttgacac cttgtagtcc cagctactca 360
tgagggcttg agctgggagt atcacttgag ccaggaatg tgcattgattt gcaatgtatc 420
caagatcatg ccaactgcact ccatgcctgt ggaaacatat gtgagatcct ggtctcaaaa 480
ataactaata cataataata cggcctgagc gctgggtttg ctctactgcc ttgtaaatcc 540
caagcacttc tgggagggcc aacttgcggt gtggatc 577

<210> 110
<211> 725
<212> DNA
<213> Homo sapien

<400> 110
tcgcgccga ggtctcgtga gccccctaga ccatcacgga tgccgagctt cgggtaactc 60
tcacagtga aggttccac gccgccccta atcccgctcg aagcagccct gagaaacatc 120
gccattctc tctccatctc acccccacaa aatttttgcc accccaacac ttcaactacta 180
tttgttttat ttttcttatt aatataagac ggcaggaatg tcaggcctct gagcccaagc 240
caagccatcg catccctgt gacttgcaag tatatgcca gatggcctga agtaactgaa 300
gaatcacaaa agaagtgaat atgctctgcc ccaccttaac tgatgacctt ccaccacaaa 360
agaagtgtaa atggccggtc cttgctttta gtgatgacat taccttgtga aagtcctttt 420
cctggctcat cctggctcaa aaatcacccc cactgagcac cttgcaacct ccaactcctgc 480
ctgccagaga gcaaaccctc tttagctgta attttcctt acctacccaa atcctataaa 540
acggccccac ccttatctcc cttegtgac tctcttttcg gactcagccc gcctgcaccc 600
aggtgaaata aacagccacg ttgctcacia aaaaaaaaca aaaagcctgg gggaaccccg 660
ggccaagcgg tcccggggtg atttggttcc ccggtcccat tccattgaa aacggtttcg 720
cacac 725

<210> 111
<211> 968
<212> DNA
<213> Homo sapien

<400> 111
 ggtcatactc ctattcaccg ttctcaacta ctcatacatg ccttgcctctt gtttacactg 60
 ccggtttaca ctgtttttcc aagccatcac agctgatata tcctgggtgct atccccaaac 120
 tgccactctt aactcttgaa gtaaataaat catctttgct ggcaggacta tgctgaatct 180
 ccttaggcac tctctaataca gacatcctga gtcgtcccaa ttcttagacc ttttatacct 240
 gtttttctcc ttctgttatt ccatttagtt tttcaattca tacaaaaccg tatccaggcc 300
 atcaccaatc attctatatg acaaagtgtt cttctaacat cccacaatc tcaccocctta 360
 ccacaagacc tcccttcagc ttaatctctc ccactctagg ttcccacgcc gcccctaatc 420
 ccgcttgaag cagccctgag aaacatcgcc cattctctct ccataccacc ccccaaaaat 480
 gttcgccgcc ccaacacttc aacactattt tgttttattt ttcttattaa tataagaagg 540
 caggaatgtc aggcctctga gcccaagcca agccatcgca tcccctgtga cttgcacgta 600
 tacgcccaga tggcctgaag taactgaaga atcacaaaag aagtgaatag ccctgccccca 660
 ccttaactga tgacattcca ccacaaaaga actgtaaatg gccgggtgctt gccttaactg 720
 atgacattac cttgtgaaag tccttttctt ggctcatcct ggctcaaaaa gcacccccac 780
 tgagcacctt gcaacccccca ctcttgcttg ccagagaaca aacccccctt gactgtaatt 840
 ttcttttacc tacccaaatc ctataaaacg gccccaccct tatcttcctt tcgtgactc 900
 tctttttgga ctcagcccg ctcacccag gtgaaataaa cagccatgtt gtcacaaaa 960
 aaaaaaaa 968

<210> 112
 <211> 535
 <212> DNA
 <213> Homo sapien

<400> 112
 tggtcgcggc gaggtaaccc ctgtaacctg tgaatattat atgacataat ggactttgca 60
 catgtaatta aattaaggat cttgagatga ggggattatc ctagattatc tgggtggacc 120
 octaaatgca atcccaagtg tccttatgag tgggggccag agagagagat tggacacagg 180
 agaaggaggc aatgtgacta ctgcagcaag atgctacact gctggcttgg aggtggaaga 240
 gaaggccaaa aatgcaacga acgtagcttg gaagctggaa aaggcaagga aactgttttc 300
 ccttagaacc tctggaggaa gtgtggccct gccaacacac tgatttttagc ccagtgaaac 360
 taattttgaa tttctgacct ctagaactgt aagagaataa atgtgtttgt ttttaagtcg 420
 aaaaaaaaaa aaaaaaaaaa aaaaggctgg gggtagcggg gccatagggg tcccgggggg 480

aattgggttc ccgccccaaa ttcccccaaa aataggggag acaatcccc aagaa 535

<210> 113

<211> 7510

<212> DNA

<213> Homo sapien

<400> 113

cagtcagtta ctcaatattt atggaacact tcctgtgcca ggccctggat ttctcagctg 60

taacacagat aaccaatatg tgccttatcc ctcaaatagc tgaccaacg gggcccaaac 120

cccattttca tggctctacc tgtcctacca actttttttt ttttttttgc caatttctgg 180

tttttcttcc tctagatcct gcctgcagac atttatattt gaacctgtgt tcttggttcc 240

agacaatgtg gggtcccacc aaaaccact tgcaccgaac tctcatgggc ttattagtaa 300

ctctgtgaaa gagttggctg cagtgggtgg ggtgaggggt ttctggacc ccttcacaca 360

ccacttagcc ctctctgact ggcccttctg ttaccactcc tccgctttgc tctgaacaag 420

tgaccctttc cctggcccag caaaccaaga gggcgtgaac aagccagtcc cgctacctgg 480

cgcttctccc agggagcatt ctccctccct tctctggccc cttctgtatt tttatgggtg 540

tttcccagg ctgctaatta attagccttc ttacaaagg cgggtgctct acctcttctt 600

cagggttggc tgtgttcatt tgtttagaac attgtccgt ctcataaatt gggtggttat 660

tagacttttt gttggtttat gactactgca tggaaatttc aggatagtca agcacattta 720

gagaaatttg gtgactgggt gaaataattg ttagggaagg atgcagatgt tctgtgttct 780

aagcaggctt gttaaccaga ttggaaatta ctgtttatat ctcatttttg ctgcatagga 840

gctcattgag ttagaatagt ttccctttct actctgcta ctgtcaaagc ctcaagatgc 900

aaaacacatt cagtttagtt gagcagctcc tgcagtctgt tcccagcacc agctatctgc 960

aaagcgccat gttagaggta agctgctaac ttccatacag cctacaaaac ccagatggcc 1020

ttccctatct cctggaagcc ttctctgact gctaccacca ctttccctt ccctagagtt 1080

aatcttttcc tttctgcaag tggcctgcgc atcctatttg ggcatttggt tcatactgtt 1140

cggttaatttt ttagcatcgg cctctcttac tagactgcgc actgaacctc tgagtccagg 1200

gactgtgtct ccttcagtgt tatatttcca gcaccagtg ggagacttta atagtcatta 1260

gtaaagtact gaatgaatga ggaaaccttc aaagccagct tacatgagtg tttcatgggg 1320

atttggtgaa cagagatgcc ctgtctcagg gtctcagtc cgtatctctc tcctgtgctg 1380

ttgcggcagc ttctgtctc tctgtgctg ctttgtctct gcagtctgtt ccagcatgc 1440

agatcccgtc acccctcttc ccggaaatct ccagtgtctt ccatcttact tcatggaaaa 1500

gctagagcat tcaattagac cccaaagatc tgactcccag ggccttcctg aaccccatca	1560
cttctgaccc ataatatata cactctacca agctaccacag aagcgctgc tccccctgga	1620
atgtactccc ttgtcaaata ttctgatcta gatgttctc tgcctagaat gttcttccca	1680
gatgcttgca tggctcactt ctttacttcc tcgtgtcttt tcccagctgc caccatttca	1740
gtgatgcctt atcaggacat catctctgcc aacagtctca cttcattctt ttctagaata	1800
cttaccacct tttaatgtac tatatatatt gacttatcca ttgttttttt gtctccctcg	1860
ctgaatgcaa gccctatgaa ggcagggatt tgtggttctt ctggttcatc atcctttccc	1920
tcacactcca gggactagat tgcattgtct catggtgaca gccaacatgc agatggtgtt	1980
tttgtgtcag gcgcttttct gagcatgtta tacattatga ttaataaacg gaggaggtaa	2040
gagtccacca ggcagtctga gctctgagcc cttatgcttt tctgcctagt agcaagtatt	2100
tgttcaaaga atgaaggaat gactacttat gataaggaga taaagcaagt tggcactgcc	2160
ttattagcct gtgttgagc tgaatgaaac agaggattcc cagccttctc cgagcaccca	2220
atgggactgg cacttatggg agagaagaag ctatgacagg ggcttcatt tcagttgtgc	2280
cctaaaagca gctgggcca aaggtggatc ttgtggatgt aggtgaatta tcatatgtaa	2340
tcagccaact caacttttac ctccagactc ttgcagcaaa gaatggtcaa ccttggaaca	2400
ttatctttaa catctcaaag ttctttctta tcagtcacct cttttcttc tcacaggaat	2460
tagataagga ggaaggaca gatggcatca tcctcathtt atgtgtgaga gcaggcagag	2520
ctgaagagac atctatgaag tcccacagct gggagccaga atcctaacc agatccccag	2580
ctttctctg tcaccttcag gcctccctct ggtcagtgtg atcatcttta gactctccag	2640
gtggcagcaa gctttgagaa gttcctgctg ctatttttac tattcctgat tctgttgacc	2700
ccaaaacag actgctgaga ggaatgctaga ccatgctcct ggggagcacc agaaaggggt	2760
cctggagctc cagtgttcag ctgcaggatt gaactaagca cacctgttga gggggacctt	2820
gagaggggag gctgctgtag ttacaaaaga aatgaacaaa tagggcatat cagccctgca	2880
gtgcagcttt aaggacatct cccatgggtt gagccttgcg gtgccatcct gccctgctgg	2940
acttgagggt caaggtgtcc attcatctag ggctgcccc gattctgcag atgtctcaa	3000
tgcagtacct tggaagctct acggggcaca ggcagctctt ccaagaagg ctaacagggt	3060
gaaaaggagc atttatggag cacctgccat ttgccagcat tctgctcagc tccatgtctt	3120
actatctgtt agcctcactc ttgcagaca agcaggagat gagaccagac tgggtaacac	3180
accaaagca ccacagtggg ggggcaaggc tcccagtctg ctcttctga cttgaagcac	3240
attccatcgc catatcactc atcacttctt ccttgtcaaa caaagcaaga ctctgcacta	3300

gcttcagget tttgttttgt tttttgacac agagtctcgc tctgtcgccc aggctggaga 3360
gcagtggcgc catctcggct cacggcaacc tctgcctccc gggttcaagc aattctcctg 3420
tcttagcctc ctgggtagct gggattacag gcgcgcgcca ctatgcctag caattttttt 3480
gtatttttag tagagatggg attttaccac attggtcagg ctggctctcaa actcctccac 3540
tcaggtgatc caaccacctt ggcctcccaa aatgccagga ttacaggcat gagccaccat 3600
gcctgcctgc tgcctattta aatagcaagc agttacagtt aaagaaagac cctggcctgg 3660
ggacgctgcc aaggctccta atctgacact tctttcatgt ggaccagggg tctgaactgt 3720
gtttcttcca aacttttgga gcttgcttcc ttggtggtaa gctaaatagt ggcccccaa 3780
atataccctt ttataacccc tgtaacctgt gaattattata tgacataatg gactttgcac 3840
atgtaattaa attaaggatc ttgagatgag gggattatcc tagattatct ggggtggacc 3900
ctaaatgcaa tccaagtgt ccttatgagt gggggccaga gagagagatt ggacacagga 3960
gaaggaggca atgtgactac tgcagcaaga tgctacactg ctggctttgg aggtggaaga 4020
gaaggccaaa aatgcaacga acgtagcttt ggaagctgga aaaggcaagg aaactgtttt 4080
cccttagaac ctctggagga agtgtgccct gccaacacac tgatttttagc ccagtgaaac 4140
taattttgaa tttctgacct ctagaacttg tcaggacatt tctagaatct gttgggggtc 4200
ctgagagaga gagagcacct gccactggca ttgaaaaaga tgtcctggcg tccgcaatac 4260
cgtagctcca agttccggaa tgtctacggg aaggtggcca accgggagca ctgcttcgat 4320
gggatcccca tcaccaagaa tgtgcacgac aaccacttct gtgccgtcaa caccogcttc 4380
ctggccatcg tcaccgagag cgcagggggc ggctccttcc tcgtcatccc cctggagcag 4440
acaggcagga ttgaacccaa ctaccccaag gtctgcggcc accagggcaa tgtgctggat 4500
atcaaatgga accccttcat cgacaacatc attgcctcgt gctcggagga cacgtcggtg 4560
cggatctggg agatccccga gggcgggctg aagcggaaca tgacggaggc gtccttgag 4620
ctgcacgggc acagccggcg tgtggggctg gtcgagtggc accccaccac caacaacatc 4680
ctgttcagcg ctggctacga ctacaaggtc ctcatctgga acctggatgt gggtagagcg 4740
gtgaagatga ttgactgcca cacggatgtg atcctctgca tgtccttcaa cacggacggc 4800
agcctgctca ccaccacgtg caaggacaag aagctgcgtg tgattgagcc ccgctctggc 4860
cgtgttctgc aggaggccaa ctgcaaaaac cacagagtga accgggtggg gttcctgggg 4920
aacatgaagc ggctcctcac gacaggggtc tccaggtgga acacaagaca gattgccctc 4980
tgggaccagg aggacctctc catgcccctg atcgaagagg aaattgatgg gctctctggc 5040

ctcctgttcc	ccttctatga	tgctgacacc	cacatgctct	acctggctgg	aaaggggat	5100
ggaaacatcc	ggtactacga	gatcagcact	gagaagccct	acctgagtta	cctcatggag	5160
ttccgctccc	cagccccgca	gaaaggccta	ggggctcatgc	ccaagcacgg	gctggatgtg	5220
tcagcctgcg	aggtgttccg	cttctacaag	ctggtgactc	tcaagggcct	gatcgagccc	5280
atctccatga	tcgtgccccg	gaggtcagat	tcctaccagg	aagacattta	cccaatgaca	5340
ccaggcacgg	agccagcact	gaccccgat	gaatggctgg	gaggcatcaa	ccgagatccc	5400
gtgctgatgt	ctttgaaaga	aggctataag	aagtcctcaa	aaatggtatt	taaggctccc	5460
atcaaagaaa	agaagagtgt	tgtggtcaac	ggaatagatt	tattagaaaa	tgtcccaccc	5520
aggacagaga	atgagctcct	tcgaatgttc	ttccggcagc	aggatgagat	tcgacggttg	5580
aaagaggagc	tggcccagaa	ggacatccgc	attcggcagc	tccagctgga	actgaaaaac	5640
ttgcgcaaca	gccccagaa	ctgttagctc	cccagctggg	ctgttttcta	agccgatctc	5700
tcctgtgttt	ctactcatcc	cttaacttct	cccttaccag	tgaccccaga	gacagagcca	5760
ggacaggagt	gggggcccagc	ctgaggaccc	ccgcctacca	cctcgagaac	tggaagccaa	5820
cctctaacct	cctgacctca	tgctaataaa	agtccccagc	ttctggagac	cccctgccgg	5880
cagccccctt	ccctgccacc	ccaggagcca	ggcttccct	cagctgggtg	aagactacag	5940
actccctggg	gttggcaggg	gtcccatctc	agtggaccag	gaagcaagag	gggaagcggg	6000
atcccagcta	gacttagaac	ttggactttt	cccctgtgaa	gggggctgcc	aggacatctc	6060
agcactcccg	cctggagctc	tcagcatcac	tgaaggatcc	acagtgtgag	tgctggactg	6120
caggctgcag	tgatccctct	ttcgtccacc	cccctcttcc	ctcagcagcc	ccggaagcct	6180
gcctcaccgg	acgaggacag	cgagcggccc	ggctccttcc	tgtctcttcc	cttcccaccc	6240
tcttgtcttc	aggggaattca	gaggattgct	ttccaaggcc	ataatgaccc	cttgccttcc	6300
ccatgattct	ctacaaagct	cttgacaccc	cttttcccat	tcaatttggtg	agccaggcag	6360
ggtagggatt	agtgtccccc	tttgacaaat	gacagaactg	agggttgcaa	tggggaaatg	6420
acttataaag	tcaccagca	ggtcaacaat	ggggccacga	ccaagaccct	gggtgttcag	6480
accccaaggc	cagggccttt	cccgtgcat	caagatgcca	atccctttgt	gggcttcacc	6540
agtgcccaag	tctctatgga	gaatgagaac	tggaagccac	tgctaccgtc	taccagcac	6600
cagtagtgcc	gatgtgccac	actgccagct	tgaggccct	cacgctctgt	gcccctagat	6660
ccttcaggct	cccacccctca	gctgtcacca	ccaccctccc	caggggactc	catctgagat	6720
gaggcctcgt	cctctctgga	gctgaggctg	agaaggggtg	agcttgggcc	tggggaaggc	6780
agaccagggt	ctgatggctt	ctagggatgc	tctgcgtgtg	tctcagcacc	gctatctcag	6840


```

ccactttcag ccttatgcac gtagaatgac cacagccact cgcacccgta tagcacttta 6900
aagttttctgc agtcctttga cacataggat ctcacgcagc ctcacgtcta ctcccttctg 6960
cagatgagga aaccgagaga agtggcccaa ggtcacgcaa ctctgagatg ccacatttca 7020
tttgatcttg tacacatttt cttttattcc ttcttttttc ctcccttcat ttcccactac 7080
gcacaaagag ttataaaca ctgttctcag aagagtcaca gtttggggtg agatctggaa 7140
atcaagaaat ggggtgtccac tcttttcttt cattagctag gatctactag atgcattata 7200
ctccatacct gctttttcca tggccgcctt acggaaaatc ccatccacag aggccagggc 7260
taccgaagcc cctccagggtg agctgggctt ttcttttatg aacctccatc ctcccagcca 7320
gctacagtag ggctctctca ccccgtagcc cacagctaga cagtgtcagc actcatctcc 7380
tcctcccaca tttctggagc tttttttttt ccttcccatc tgacctttgt ggtcttctgt 7440
gattatttat gctgcctccc aaggatagaa ttgaaataaa atgttttcaa cttagaaaaa 7500
aaaaaaaagg 7510

```

```

<210> 114
<211> 917
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (616)..(616)
<223> a, c, g or t

```

```

<400> 114
gaaaaagaag atgatcatat gggcgaatgg gccctagatg ctgctcgagc ggccgaggtg 60
tgatggatga gcggccgccc gggcagggtg taatcccagc tacttgggag gctgaggcag 120
agaattgctt gaaccggga ggcagagggt gcagtgagtc gagatcgta cactgcactc 180
cagccaggca acagaaggag actccatctc aaaaaaaga aaaaaaggta aggccggact 240
cagtggctca cacttgtaat ctcagcactt cgggaggagg ctgaggcagg cagattgctt 300
gcgcttagga gttcaggact gaactaggca acatggagaa accatgtctc tacaaaatat 360
aaaaaaatta gctggacatg gtgtcttgca cctgtagtcc cagctactca ggaggctgag 420
ctgggagtat cacttgagcc caggaagtgc agattgcagt agccaagatc atgccactgc 480
actccagcct gggaaacata gtgagatcct gtctcaaaaa taataataat aaaataggcc 540
gagcgcggtg gtcacgcct gtaatccag cactttggga ggccaaggcg ggtggatcac 600
gaggtcagga gatcangacc atcctggcta acacggtgaa accccatctc tactaaaaat 660

```

acaaaaaatt agccccggtgt ggtggtgggc gcctgtagtc ccagctacta gggaggcgga 720
 gggaaggaga atggcgtgaa cccgggaggt ggagcttgca gtgagccgag attgcaccaa 780
 tgcactcagc ctgggtaata cagcgagact ccatccaag aaaaaaaaaa aaaaaagag 840
 cgtggggaca ctgggcaaag ggccgtggaa tggtcgcca caaaaacaa caaaaagaaa 900
 aaaaaaagc ccaaaaa 917

<210> 115
 <211> 787
 <212> DNA
 <213> Homo sapien

<400> 115
 ccgccccgggc aggttgtaat cccagctact tgggaggctg aggcagagaa ttgcttgaaac 60
 ccgggaggca gaggttcag tgagtcgaga tcgtaccact gcaactccagc cagggaaca 120
 gaaggagact ccatctcaa aaaaagaaaa aaaggttaagg ccggactcag tggctcacac 180
 ttgtaatctc agcacttcgg gaggaggctg aggcaggcag attgcttgcg cttaggagtt 240
 caggactgaa ctaggcaaca tggagaaacc atgtctctac aaaatataaa aaaattagct 300
 ggacatgggtg tcttgacact gtagtcccag ctactcagga ggctgagctg ggagtatcac 360
 ttgagcccag gaagtgcaga ttgcagtagc caagatcatg ccaactgcact ccagcctggg 420
 aaacatagtg agatcctgtc tcaaaaaataa taataataaa ataggccgag cgcggtggct 480
 cagcctgta atcccagcac tttgggaggc caaggcgggt ggatcacgag gtcaggagat 540
 caagaccatc ctggctaaca cggtgaaacc ccatctctac taaaaataca aaaaattagc 600
 ccggtgtggt ggtgggcgcc ttagtccca gctactaggg aggcggaggc aggagatggc 660
 gtgaacccgg gaggtggagc ttgcagtga gccgagattg gaccactgca ctccagcctg 720
 ggtatacagc gagactcatc ccaaaaaaaa aaaaaaaaag ctggggtaac ctggcaaacc 780
 gtcccgg 787

<210> 116
 <211> 666
 <212> DNA
 <213> Homo sapien

<400> 116
 taatgcatgc tcgagcggcg ccagtgtgat ggatgcgtgg tcgcggccgg aggtaccatc 60
 aaaacggaag gcgagcatga ccctgtgacg gagtttatag gtgaggccga ctgcctagcc 120
 ctttactaca atagaaaatg tcagctgggc gcggtggctc atgcctgtaa tcccagcact 180

82

```

ttgggaggcc aaggtgggtg gatcacctga ggttgggagt tcaagaccag cccgaccaac 240
atggagaaac cccatctcta ctaaaaatac agcattagct gggcatggtg gcacgtgcct 300
gtagtcccag ctactcagga ggctgaggtg gggagaatcg cttgaacctg ggaggcagag 360
gtttcagtga accaagatct catgccacgg cgctccagct tgggtgacaa agtgagactt 420
tatctcaaat aaataaataa ataaataaag tcagggtgtg gtggctcacg cctgtaatcc 480
cagcactttg ggaggcagag gcaggtgggt cacgaggtca ggagttcaag accagcccga 540
ccaagatggt gaaatcccg tctacaaaa aaaaaaaaaa aaaaaaaaaa ggttgggggt 600
accggggcca aaggggtccc tgtgtgaatt tgtttccgc tccatttccc cacatttgaa 660
aaaacc 666

```

```

<210> 117
<211> 664
<212> DNA
<213> Homo sapien

```

```

<400> 117
tgatgatcat atggggcatg ggtctctaga tgcattgctg agcggcgag tgtgatggat 60
gggccgcccc ggcaggtacc tgctggagaa gggagactac aaggacagca gcgagtttgg 120
ggcccgtcac cctcaggac atggatgaag ctggaaacca tcattctcag caaactaaca 180
cagaagcaga aaaccaaaca ccacatgttc tcactcataa gtgggagttg aacagtgaga 240
acacatggac acaaggaggg gaacatcaca caccgaggcc tgtcaggag tgggggacaa 300
gaggagaga gagcatggga caaatccaa atgcatgcac ggcttaaaac ctagatgaca 360
ggttgatggg tgcagcaaac caccatggca catgtatacc tgtgtaaaa acctgcacgt 420
tctgcacatg tatcccagaa cttataataa aaaagaaacc aggaaaaaag aaaaaaaca 480
aaacaaaaaa aacggtcggg gcgtcatcac ggggtccata agctggtctc cgggggtgga 540
ctatgggttt ttcccgtcc acatattccc ccgacaacta acgaggagg caacgggaca 600
agcgaacaag agagaaagcg agcaagagag tacacaaagc agagccagcc agagcaagag 660
gcag 664

```

```

<210> 118
<211> 708
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (185)..(185)
<223> a, c, g or t

```

<220>
 <221> misc_feature
 <222> (521)..(521)
 <223> a, c, g or t

<400> 118
 attcgtggtg acgatcttct atagtctggc gtgactggct ctactgcaga gtactgggaa 60
 ctctgagggg aatcgtcacg taggtctaca tcggagatta ggacaagctc tgttgatac 120
 aaccattagg ctatacctta tggctgtgac ttagtaaaca gcgagacgca cgtatggctt 180
 tgcgnaatga gaaccattct tgaatcttac ggtgggctcg ttgaagcgat agagaaggca 240
 tgtgatccta ggatcatggg gagcaccac cgagtataag ggccacgcac aactgacgc 300
 ctcgtagcat ctgcacgac atgaacacca cgcttcgcgc gagttactct aagagatcat 360
 gtcgaatacg ctttgattct cgtcacagga gcacccatca ggcacacggc atttgggagg 420
 tacactcact aggtcctac gtctttgctt cctctagtgc tgcaatactt gcttctcccg 480
 ggtctatcaa ttctgcatc aaatgatggg ggaggcagaa ntgaaactaa acttaaagg 540
 ggaatccttt gcagaaaaaa aaacaaagaa gaaataaaaa aaaaagagtg tgtggcggta 600
 accgtgtggg ccaatgaggg ggtgttcgc gtgtgtgtgt ggaaagtgtg gttttctcgc 660
 gcccaaaatt tccacccaaa caatttcggg agaacaacag gggaagga 708

<210> 119
 <211> 36
 <212> PRT
 <213> Homo sapien

<400> 119

Met Pro Val Glu Asn His Glu Leu Arg His Ile Leu Pro Gln Phe Glu
 1 5 10 15

Glu Lys Ile His Lys Lys Trp Arg Thr Thr Ile Leu Gly Ser His Ser
 20 25 30

Thr Phe Arg Glu
 35

<210> 120
 <211> 40
 <212> PRT
 <213> Homo sapien

<400> 120

84

Met Arg Ser Trp Thr Lys Asp Ile Tyr Ser Phe Ile Gln Tyr Ser Cys
 1 5 10 15

Val Cys Val Leu Glu Thr Gly Ser Cys Ser Val Gly Gln Ile Gly Leu
 20 25 30

Asp Val Ser Pro Leu Ile Asp Gln
 35 40

<210> 121
 <211> 185
 <212> PRT
 <213> Homo sapien
 <400> 121

Met Ile Thr Thr Ser Thr His Ile Tyr Pro Leu Thr Thr Leu Leu Thr
 1 5 10 15

Asp Ser Arg Thr Leu Leu Ile Arg Leu Tyr Ile Leu Asn Arg Phe Thr
 20 25 30

Pro Ala Ser His Ser Ile Gln Pro Thr Gln Leu Pro Pro His Pro Leu
 35 40 45

Ile Ser His Ser Leu Leu Pro Thr His Asp Pro Leu Pro Ser Thr Ser
 50 55 60

His His Ser Thr Thr Gln His Leu Leu Pro Leu Ile Thr Thr Pro Ser
 65 70 75 80

Thr His Ser Pro Pro Ile His Tyr His His Pro Pro Asn Pro Pro Leu
 85 90 95

Pro His Met His Thr Ser Pro His Ser Pro Thr Tyr Asn His Leu Ser
 100 105 110

His Ile Pro Leu Asn Gln Pro Pro His His His Arg Leu Asp Ser Ser
 115 120 125

Ser Pro Thr His Pro Pro Leu His Ile His Lys Gln Ile Asn His Thr
 130 135 140

Ser Ala Pro His Asn Thr His Thr Arg Ser Thr Leu Thr Pro Pro Pro
 145 150 155 160

85

Pro Thr Leu His Ser His Ser Ser His Ser Pro Leu Thr Thr Pro His
 165 170 175

His His Leu Leu Ser Pro Leu Pro Pro
 180 185

<210> 122
 <211> 36
 <212> PRT
 <213> Homo sapien

<400> 122

Met Arg Asp Ser Asn Leu Asp Pro Gly Thr Ser Lys Tyr Val Ser His
 1 5 10 15

Val Ser Leu Trp Trp Val Pro Pro Ser Leu Asn Gly Gly Cys Cys Leu
 20 25 30

Gln Val Asn Asn
 35

<210> 123
 <211> 76
 <212> PRT
 <213> Homo sapien

<400> 123

Met Gln Gly Leu Leu Ile Pro Val Ser Cys Ser Ile Thr Val Thr Leu
 1 5 10 15

Cys Pro Phe Phe Pro Pro His Asn Phe Tyr Phe His Asn Phe Leu Phe
 20 25 30

Val Ser Ile Leu Phe Leu Lys Ser Leu Ser Phe Ser Ile Gly Leu Phe
 35 40 45

Leu Ser Val Ser Asn Cys Val Ser Leu Leu Ser Val Cys Leu Cys Ile
 50 55 60

Ser Leu Pro Ile Ser Ala Tyr Leu Phe Phe Ser Phe
 65 70 75

<210> 124
 <211> 23
 <212> PRT
 <213> Homo sapien

<400> 124

Met Arg Leu Ala Pro Trp Tyr His Leu Leu Pro Glu Ile Phe Pro Phe
1 5 10 15

Ser Thr Arg Ala Lys Val Leu
20

<210> 125

<211> 70

<212> PRT

<213> Homo sapien

<400> 125

Met Ala Met Val Ala Met Gln Pro Val Asn Leu His Ala Ile Phe Trp
1 5 10 15

Glu Gly Leu Arg Val Gly Gly Ile Ala Leu Thr Ala Ala Gly Trp Lys
20 25 30

Val Ala Ser Glu Val Lys Glu Thr Gln Ala Ile Gln Val Arg Gly Gln
35 40 45

Glu Gln Asp Ser Ile Ser Lys Lys Lys Lys Lys Lys Lys Glu Glu Pro
50 55 60

Val Pro Arg Pro Arg Pro
65 70

<210> 126

<211> 104

<212> PRT

<213> Homo sapien

<400> 126

Met His Phe Lys Gly Gln Gly Ala Gly Gly Leu Thr Pro Val Ile Pro
1 5 10 15

Ser Thr Leu Gly Arg Ala Glu Ala Gly Gln Ile Thr Arg Ser Gly Asp
20 25 30

Leu Arg Pro Phe Leu Gly Leu Thr Arg Val Lys Pro Leu Ser Leu Leu
35 40 45

Lys Ile Gln Lys Lys Lys Phe Ser Arg Gly Val Val Gly Gly Ala Pro
50 55 60

Cys Leu Ser Gln Ala Tyr Ser Arg Gly Leu Arg Ala Gly Asp Trp Ala
65 70 75 80

Asp Pro Gly Gly Arg Asp Ala Leu Leu Leu Ser Gly Asp Ser Arg Leu
85 90 95

Gly Phe Gln Ala Trp Ala Arg Trp
100

<210> 127
<211> 23
<212> PRT
<213> Homo sapien

<400> 127

Met His Val Glu Arg Pro Gln Phe Val Met Asp Pro Thr Leu Gln His
1 5 10 15

Tyr Leu Phe Tyr Phe Ser Tyr
20

<210> 128
<211> 17
<212> PRT
<213> Homo sapien

<400> 128

Met Pro Leu Asp Phe Ser Pro Gly Asp Pro Ser Trp Thr Ser Asp Pro
1 5 10 15

Gln

<210> 129
<211> 16
<212> PRT
<213> Homo sapien

<400> 129

Met Gly Ser Ile Val Asn Phe Thr Lys Lys Ala Lys Leu Cys Lys Tyr
1 5 10 15

<210> 130
<211> 19
<212> PRT
<213> Homo sapien

88

<400> 130

Met Ile Lys Thr Ser Lys Ala Asn Gly Asn Glu Asn Lys Asn Arg Gln
 1 5 10 15

Ile Glu Thr

<210> 131

<211> 61

<212> PRT

<213> Homo sapien

<400> 131

Met Glu Gly Arg Ala Leu Leu Glu Ser Leu Leu Ala Leu Ser Cys Val
 1 5 10 15

Gly Ala Gln Val Pro Leu Ser His Pro Pro Arg Gly Asp Leu Gly Ser
 20 25 30

Gln Pro Pro Ile Ile Pro Pro Pro Trp Gly Glu Ser Leu Ala His Pro
 35 40 45

Gln Ala Phe Lys Lys Cys Pro Leu Ile Gln Arg Lys Lys
 50 55 60

<210> 132

<211> 29

<212> PRT

<213> Homo sapien

<400> 132

Met Pro Ser Gly Gly Ile Cys Asp Gly Leu Val Ala Ala Arg Tyr Tyr
 1 5 10 15

Thr Leu Leu Val Thr Ile Val Leu Tyr Asn Ser Lys Phe
 20 25

<210> 133

<211> 32

<212> PRT

<213> Homo sapien

<400> 133

Met Trp Gln Asn Pro Val Ser Thr Lys Ile Gln Ile Leu Leu Gly Leu
 1 5 10 15

89

Trp Ala Ala Leu Val Ser Gln Leu Leu Arg Gly Trp Glu Glu Ile Ala
 20 25 30

<210> 134
 <211> 39
 <212> PRT
 <213> Homo sapien

<400> 134

Met His Ala Glu Arg Arg Ser Val Met Asp Gly Arg Pro Gly Arg Tyr
 1 5 10 15

Trp Asp Tyr Arg His Glu Ser Arg Cys Leu Ala Phe Ser Gln Ile Phe
 20 25 30

Lys Ser Arg Val His Gly Ser
 35

<210> 135
 <211> 94
 <212> PRT
 <213> Homo sapien

<400> 135

Gln Ser Leu Thr Leu Ser Pro Arg Leu Glu Cys Ser Gly Thr Val Ser
 1 5 10 15

Ala His Cys Asn Leu His Leu Leu Gly Ser Ser Asp Ser Pro Ala Ser
 20 25 30

Val Ser Ala Val Ala Gly Thr Thr Gly Val Arg His His Ala Trp Leu
 35 40 45

Ile Phe Ile Phe Leu Val Glu Thr Val Phe Cys His Val Gly Gln Ala
 50 55 60

Gly Leu Lys Leu Leu Thr Ser Gly Asp Pro Pro Thr Ser Ala Ser Ala
 65 70 75 80

Ser Thr Gly Ile Thr Gly Met Ser His Cys Ala Trp Pro Ser
 85 90

<210> 136
 <211> 55
 <212> PRT
 <213> Homo sapien

90

<400> 136

Met Cys Met Ser Ala Asn Leu Gly Tyr Pro Gly Ala Ala Thr Gly Ala
 1 5 10 15

Arg Tyr Arg Thr Val His Lys Asn Leu Ser Val Pro Ala Leu Lys Lys
 20 25 30

Pro Thr Cys Pro Pro Val Asn Leu Pro Gly Thr Val Leu Gly Cys Glu
 35 40 45

Gly Met Glu Thr Thr Lys Ala
 50 55

<210> 137

<211> 76

<212> PRT

<213> Homo sapien

<400> 137

Met His Met Cys Lys Lys Asn Ser Thr His Leu Lys Asn Lys Asn Asn
 1 5 10 15

Lys Gln Lys Glu Lys Lys Arg Ala Leu Trp Gly Cys Thr Pro Val Gly
 20 25 30

Gln Lys Arg Val Cys Pro Pro Trp Cys Val Ser Asn Phe Val Phe Ser
 35 40 45

Pro Arg Pro Pro Ile Phe Pro Pro Lys Ile Ile Arg Glu Lys His Lys
 50 55 60

Gly Gly Trp Thr Leu Ala His His Thr Leu Ile Ala
 65 70 75

<210> 138

<211> 69

<212> PRT

<213> Homo sapien

<400> 138

Met Val Thr Leu Glu Arg His Thr Gly Gln Asp Val Leu Val Ala Phe
 1 5 10 15

Pro Gln Asp Pro Trp Ser Ser His Val Ala Ser Asn Leu Trp Asp His
 20 25 30

His His His Leu Ser Ser Arg Ser Leu Lys Ser His Ala His Leu Ser
 35 40 45

Phe Arg Gln Ser Ser Phe Cys Glu Val Cys Ile Ser Ser Leu Thr Thr
 50 55 60

Leu Cys Arg Ser Phe
 65

<210> 139
 <211> 39
 <212> PRT
 <213> Homo sapien

<400> 139

Met Arg Asp Gly Lys Arg Glu Thr Ala Lys Arg Gly Glu Arg Val Ser
 1 5 10 15

Glu Phe Gly Lys Gly Leu Lys Ala Gln Ala Gly Cys Leu Lys Pro Phe
 20 25 30

Lys Pro Pro Val Leu Ser Pro
 35

<210> 140
 <211> 332
 <212> PRT
 <213> Homo sapien

<400> 140

Met Thr Thr Ser Leu Asp Thr Val Glu Thr Phe Gly Thr Thr Ser Tyr
 1 5 10 15

Tyr Asp Asp Val Gly Leu Leu Cys Glu Lys Ala Asp Thr Arg Ala Leu
 20 25 30

Met Ala Gln Phe Val Pro Pro Leu Tyr Ser Leu Val Phe Thr Val Gly
 35 40 45

Leu Leu Gly Asn Val Val Val Val Met Ile Leu Ile Lys Tyr Arg Arg
 50 55 60

Leu Arg Ile Met Thr Asn Ile Tyr Leu Leu Asn Leu Ala Ile Ser Asp
 65 70 75 80

Leu Leu Phe Leu Val Thr Leu Pro Phe Trp Ile His Tyr Val Arg Gly
85 90 95

His Asn Trp Val Phe Gly His Gly Met Cys Lys Leu Leu Ser Gly Phe
100 105 110

Tyr His Thr Gly Leu Tyr Ser Glu Ile Phe Phe Ile Ile Leu Leu Thr
115 120 125

Ile Asp Arg Tyr Leu Ala Ile Val His Ala Val Phe Ala Leu Arg Ala
130 135 140

Arg Thr Val Thr Phe Gly Val Ile Thr Ser Ile Val Thr Trp Gly Leu
145 150 155 160

Ala Val Leu Ala Ala Leu Pro Glu Phe Ile Phe Tyr Glu Thr Glu Glu
165 170 175

Leu Phe Glu Glu Thr Leu Cys Ser Ala Leu Tyr Pro Glu Asp Thr Val
180 185 190

Tyr Ser Trp Arg His Phe His Thr Leu Arg Met Thr Ile Phe Cys Leu
195 200 205

Val Leu Pro Leu Leu Val Met Ala Ile Cys Tyr Thr Gly Ile Ile Lys
210 215 220

Thr Leu Leu Arg Cys Pro Ser Lys Lys Lys Tyr Lys Ala Ile Arg Leu
225 230 235 240

Ile Phe Val Ile Met Ala Val Phe Phe Ile Phe Trp Thr Pro Tyr Asn
245 250 255

Val Ala Ile Leu Leu Ser Ser Tyr Gln Ser Ile Leu Phe Gly Asn Asp
260 265 270

Cys Glu Arg Ser Lys His Leu Asp Leu Val Met Leu Val Thr Glu Val
275 280 285

Ile Ala Tyr Ser His Cys Cys Met Asn Pro Val Ile Tyr Ala Phe Val
290 295 300

Gly Glu Arg Phe Arg Lys Tyr Leu Arg His Phe Phe His Arg His Leu
305 310 315 320

Leu Met His Leu Gly Arg Tyr Ile Pro Phe Leu Pro
 325 330

<210> 141
 <211> 57
 <212> PRT
 <213> Homo sapien

<400> 141

Met Asp Ser Gln Ser Ser Ser Pro Ala Ser Ile Ser Asp Pro Gly Gly
 1 5 10 15

Ser Pro Pro Arg Cys Val Gln Pro Ser Gly Asn His Glu Leu Ser Cys
 20 25 30

Pro Leu Gly Gln Met Leu Cys Gln Val Leu Val Leu Arg Ala Ser Pro
 35 40 45

Leu Pro Gly His Gly Gly Ala Cys Leu
 50 55

<210> 142
 <211> 56
 <212> PRT
 <213> Homo sapien

<400> 142

Met Ser Phe Lys Ser Leu Lys Phe Ser Leu His Leu Phe Phe Ser Leu
 1 5 10 15

Val Thr Tyr Leu Trp Lys Gly Ser Gly Cys Leu Ser Tyr Ser Val Pro
 20 25 30

His Cys Gln Asp Phe Glu Gly Tyr Ile Leu His Ser Ile Val Leu Arg
 35 40 45

Leu Leu Leu Trp Phe Leu Phe Leu
 50 55

<210> 143
 <211> 77
 <212> PRT
 <213> Homo sapien

<400> 143

94

Met Gln Cys Leu Arg Phe Thr Gln Val Ser Leu Leu Phe Leu Gly Pro
 1 5 10 15

Thr Ile Pro Ser Ser Ile Thr Ile Thr Ile Pro Gln Gln Thr Pro Ile
 20 25 30

Cys Tyr Pro Val Met Thr Ala Asp Pro His Asp Pro Ser Pro Cys Val
 35 40 45

Arg Gly Pro Thr Ser Ser Ile Leu Ser Ile Leu Gly Ser Asn Phe Asn
 50 55 60

Met Ile Leu Lys Gly Gln Tyr Ser Thr Ile Leu Thr Tyr
 65 70 75

<210> 144
 <211> 53
 <212> PRT
 <213> Homo sapien

<400> 144

Met Thr Ala Met Gly Thr Trp Leu Arg Trp Pro His Ser Thr Glu Ser
 1 5 10 15

Gln Gly Leu Gln Ser Gln Lys Gln Gln Gly Gln Phe Gln Ile Pro Ala
 20 25 30

Pro Leu Phe Thr Val Cys Val Thr Leu Ser Lys Thr Leu Pro Pro Asn
 35 40 45

Leu Phe Ser Tyr Leu
 50

<210> 145
 <211> 130
 <212> PRT
 <213> Homo sapien

<400> 145

Met Ser Phe Ser Phe Cys Leu Phe Leu Phe Cys Met Gly Cys Leu Val
 1 5 10 15

Leu Phe Ala Phe Ser Leu Phe Cys Ser Val Phe Cys Arg Leu Ala Ser
 20 25 30

Ser Cys Ala Arg Phe Arg Phe Arg Ala Phe Leu Trp Arg Phe Val Ser

95

35

40

45

Arg Ser Ala Leu Gly Arg Leu Arg Leu Cys Ser Cys Ser Gly Ser Phe
 50 55 60

Val Phe Leu Cys Phe Arg Val Ala Arg Ser Phe Ser Gly Pro Leu Cys
 65 70 75 80

Cys Ser Val Cys Ser Leu Ala Leu Val Val Ala Val Gly Val Leu Phe
 85 90 95

Val Ser Leu Arg Val Pro Cys Leu Cys Ala Ser Pro Val Ser Leu Leu
 100 105 110

Val Phe Arg Ser Ala Val Val Arg Ala Met Val Trp Gly Asn Cys Gly
 115 120 125

Ala Glu
 130

<210> 146
 <211> 120
 <212> PRT
 <213> Homo sapien

<400> 146

Met Glu Met Ser Pro Asn Ile Ser Pro Asn Arg Ala Pro Asn Glu Lys
 1 5 10 15

Gly Gly Gln Lys Asn Gly Ser Pro Lys Ala Ala Thr Arg Leu Ala Pro
 20 25 30

Gly Gln Ile Ala Ile Asn Asn His Leu Lys Val Ala Gln Ala Trp Trp
 35 40 45

Leu Pro Pro Arg Lys Ser Gln His Phe Gly Glu Ala Glu Asp Arg Trp
 50 55 60

Ile Ile Ser Arg Val Thr Glu Cys Arg Asp His Ala Trp Pro His Asn
 65 70 75 80

Gly Glu Asn Pro Thr Leu Tyr Leu Glu Asn Thr Gln Lys Leu Ala Arg
 85 90 95

Ala Val Asn Gly Gly Leu Pro Val Asn Pro Ser Val Leu Pro Arg Gly

96

100

105

110

Leu Arg Gln Arg Lys Lys Leu Leu
 115 120

<210> 147
 <211> 49
 <212> PRT
 <213> Homo sapien

<400> 147

Met Gly Gly Ser Gly Ser Ser Thr Pro Leu Phe Pro Cys Gln Leu Phe
 1 5 10 15

Gly Ala Thr His Ser Ser His Cys Pro Val Asn Gln Pro His Ser Leu
 20 25 30

Val Cys Trp Val Arg Arg Ser Gln Leu Glu Asp Gln Gly Leu His Tyr
 35 40 45

Cys

<210> 148
 <211> 95
 <212> PRT
 <213> Homo sapien

<400> 148

Met Cys Pro Ser Asp Tyr Pro Pro Pro Ala Val Cys Leu Phe Leu Leu
 1 5 10 15

Phe Leu Leu Trp Phe Pro Val Phe Phe Arg Val Leu Ile Pro Phe Lys
 20 25 30

Trp Lys Ser Trp Val Lys Trp Gly His Pro Pro Ala Pro Pro Arg Ala
 35 40 45

Pro Leu Pro Gln Ile Cys Pro Gln Pro Leu Gly Thr Tyr Gly Gly His
 50 55 60

Gly Gln Pro Cys Gly Ser Gln Pro Leu Pro Glu Gly Ser Ile Cys Ser
 65 70 75 80

Gly Leu Gly Glu Ala Phe Lys Ser Val Asn Leu Gln Met Leu Val
 85 90 95

<210> 149
 <211> 60
 <212> PRT
 <213> Homo sapien

<400> 149

Met Gly Thr Ala Lys Ser Tyr Lys Gly His Trp Arg Pro Gly Cys Trp
 1 5 10 15

Trp Leu Met Pro Val Ile His Asn Gln His Pro Leu Gly Glu Gly Ala
 20 25 30

Pro Asn Gly Arg Cys Ala Leu Arg Ile Pro Asn Pro Leu Met Val Gln
 35 40 45

Glu Thr Phe Asn Glu Leu Cys Leu Gly Gln His Gly
 50 55 60

<210> 150
 <211> 68
 <212> PRT
 <213> Homo sapien

<400> 150

Met Thr Thr Val Arg Gly Arg Gly Arg Ala Pro Pro Gly Ser Cys His
 1 5 10 15

Leu Ser Pro Val Arg Gly Gln Arg Val Lys Asp Leu Leu Gly Trp Arg
 20 25 30

Pro Glu Cys His Pro Glu Pro Glu Lys Phe Pro Tyr Leu Asn Glu Pro
 35 40 45

Phe Gly Phe Glu Glu Arg Ser Leu Leu Thr Pro Arg Thr Gly Leu Arg
 50 55 60

Arg Leu Phe Leu
 65

<210> 151
 <211> 61
 <212> PRT
 <213> Homo sapien

<400> 151

98

Met Thr His Met Gly Thr Gly His Leu Met Leu Leu Glu Arg Arg Ser
 1 5 10 15

Val Met Asp Trp Ser Arg Arg Gly Thr Ile Thr Gly Ser Leu Lys Pro
 20 25 30

Gln Phe Leu Ser Ser Arg Glu Pro Pro Cys Leu Ser Leu Tyr His Gln
 35 40 45

Ser Arg Leu Leu Gly Tyr Gly Leu Arg Val Leu Arg Leu
 50 55 60

<210> 152

<211> 36

<212> PRT

<213> Homo sapien

<400> 152

Met Glu Gln Val Asn Gly Lys Leu Asp Glu Leu Met Arg Val Lys Thr
 1 5 10 15

Val Glu Val Arg Asn Ser Lys Arg Arg Thr Lys Ser Ile Ala Asp Lys
 20 25 30

Lys Gln Asn Glu
 35

<210> 153

<211> 80

<212> PRT

<213> Homo sapien

<400> 153

Met Gly Phe His Arg Val Ser Gln Asp Gly Leu Asp Leu Arg Pro Arg
 1 5 10 15

Asp Pro Pro Ala Leu Ala Ser Gln Ser Ala Gly Ile Thr Gly Val Ser
 20 25 30

His Arg Ala Arg Pro Ile Leu Leu Leu Phe Leu Arg Gln Asp Leu
 35 40 45

Thr Met Phe Pro Arg Leu Glu Cys Ser Gly Met Ile Leu Ala Thr Ala
 50 55 60

Ile Cys Thr Ser Trp Ala Gln Val Ile Leu Pro Ala Gln Pro Pro Glu

99

65

70

75

80

<210> 154
 <211> 109
 <212> PRT
 <213> Homo sapien

<400> 154

Phe Phe Phe Phe Leu Gly Trp Ser Leu Ala Val Leu Pro Arg Leu Glu
 1 5 10 15

Cys Ser Gly Ala Ile Ser Ala His Cys Lys Leu His Leu Arg Gly Ser
 20 25 30

Arg His Ser Pro Ala Ser Ala Ser Leu Val Ala Gly Thr Thr Gly Ala
 35 40 45

His His His Thr Gly Leu Ile Phe Val Phe Leu Val Glu Met Gly Phe
 50 55 60

His Arg Val Ser Gln Asp Gly Leu Asp Leu Arg Pro Arg Asp Pro Pro
 65 70 75 80

Ala Leu Ala Ser Gln Ser Ala Gly Ile Thr Gly Val Ser His Arg Ala
 85 90 95

Arg Pro Ile Leu Leu Leu Leu Phe Leu Arg Gln Asp Leu
 100 105

<210> 155
 <211> 87
 <212> PRT
 <213> Homo sapien

<400> 155

Met Arg Pro Gly Ala Arg Gly Trp Pro Ser Ala Pro Val Val Ile Ser
 1 5 10 15

Pro Ser Thr Leu Gly Glu Arg Pro Arg Gly Arg Gly Gly Thr Pro Arg
 20 25 30

Arg Val Ser Gly Glu Asn Trp Glu Asn His Leu Arg Val Ala Ile Thr
 35 40 45

Thr Gly Val Lys Thr Leu Cys Val Pro Ile Leu Lys Lys Leu Pro Lys
 50 55 60

100

Lys Asn Lys Phe Arg Pro Gly Ala Val Trp Gly Ala Gly Ala Pro Val
 65 70 75 80

Phe Cys Ser Pro Glu Leu Thr
 85

<210> 156
 <211> 79
 <212> PRT
 <213> Homo sapien

<400> 156

Met Gly Phe Ser Pro Ile Phe Phe Phe Pro Pro Cys Phe Phe Leu Phe
 1 5 10 15

Ser Phe Leu Phe Phe Cys Val Glu Asn Ser Trp Gly Asp Leu Ser Pro
 20 25 30

Leu Arg Ala Leu Phe Phe Ser Arg Leu Phe Val Thr Ser Pro Val Phe
 35 40 45

Cys Ala Phe Ser Tyr Phe Leu Arg Ser Phe Ile Leu Thr Leu Val Phe
 50 55 60

Ser Leu Leu Phe Ile Phe Ser His Met Val Val Phe Leu Leu Leu
 65 70 75

<210> 157
 <211> 146
 <212> PRT
 <213> Homo sapien

<400> 157

Met Arg Cys Leu Arg Pro Cys His Ala Thr Cys Ser Val Cys Pro Ala
 1 5 10 15

Val Ser Pro Leu Phe Cys Ser Cys Leu Cys Arg Leu Ser Leu Ala Pro
 20 25 30

Pro Ile Leu Ser Trp Ser Ala Leu Pro Pro Ala Pro Ala Cys Leu Leu
 35 40 45

Phe Ser Arg Gly Pro Pro Ser Gly Ser His Ala Pro Ala Leu Pro Phe
 50 55 60

101

Cys Val Phe Val Ser Leu Cys Phe Phe Ser Phe Leu Ser Pro Leu Phe
65 70 75 80

Ser Pro Ser Ala Leu Val Cys Val Ala Leu Val Leu Ala Leu Ser Leu
85 90 95

Ala Leu Gly Ala Ala Arg Ala Pro Pro Cys Pro Gly Pro Asp Gly Leu
100 105 110

Ala Arg Val Val Val Pro Gly Val Ala Gly Gly Ala Trp Val Val Phe
115 120 125

Ser Ala Trp Leu Pro Val Trp Phe Val Val Gly Lys Leu Gly Gly Ala
130 135 140

Gly Glu

145

<210> 158

<211> 33

<212> PRT

<213> Homo sapien

<400> 158

Met Lys Glu Asp Ser His Gly Thr Leu Gly Gln Ala Arg Asn Pro Thr
1 5 10 15

Gln Leu Trp Glu Ala Glu Ala Lys Val Glu Ser Pro Arg Gly His Gly
20 25 30

Val

<210> 159

<211> 70

<212> PRT

<213> Homo sapien

<400> 159

Phe Phe Phe Phe Leu Glu Met Glu Ser Cys Ser Val Ala Glu Ala Gly
1 5 10 15

Val His Ala Ser Leu Leu Thr Glu Pro Pro Pro Ala Gly Ser Ser Asn
20 25 30

102

Ser Pro Thr Ser Ala Ser Arg Val Ala Gly Ile Thr Gly Ala Cys His
 35 40 45

His Ala Gly Leu Ile Leu Val Tyr Ala Ala Arg Gly Gly Phe His Leu
 50 55 60

Glu Thr Gly Ser His Met
 65 70

<210> 160
 <211> 84
 <212> PRT
 <213> Homo sapien

<400> 160

Met Val Ser Pro Pro Pro Phe Phe Phe Phe Phe Phe Phe Leu Tyr Val
 1 5 10 15

Phe Phe Trp Arg Gln Ser Phe Ala Leu Val Ile Gln Val Gly Val Gln
 20 25 30

Trp His Asn Phe Gly Ser Leu Gln Pro Pro Pro Pro Arg Phe Lys Gln
 35 40 45

Phe Ser Cys Leu Ser Leu Leu Ser Ser Trp Asp Tyr Arg His Thr Pro
 50 55 60

Pro His Pro Ala Asn Phe Cys Ile Phe Ser Arg Asp Gly Val Ser Pro
 65 70 75 80

Cys Trp Pro Asp

<210> 161
 <211> 116
 <212> PRT
 <213> Homo sapien

<400> 161

Pro Phe Val Pro Met Val Ser Pro Pro Pro Phe Phe Phe Phe Phe Phe
 1 5 10 15

Phe Leu Tyr Val Phe Phe Trp Arg Gln Ser Phe Ala Leu Val Ile Gln
 20 25 30

Val Gly Val Gln Trp His Asn Phe Gly Ser Leu Gln Pro Pro Pro Pro

103

35

40

45

Arg Phe Lys Gln Phe Ser Cys Leu Ser Leu Leu Ser Ser Trp Asp Tyr
 50 55 60

Arg His Thr Pro Pro His Pro Ala Ile Phe Val Phe Leu Val Glu Thr
 65 70 75 80

Gly Phe His His Val Gly Gln Thr Ser Leu Thr Pro Thr Gln Val Thr
 85 90 95

Cys Ser Pro Met Leu Cys Gly Asp Cys Glu Ala Asp Asn Asn Ile Ala
 100 105 110

His Arg Gln Gln
 115

<210> 162
 <211> 56
 <212> PRT
 <213> Homo sapien

<400> 162

Met Trp Gly Gly Val Thr Arg Trp Trp Glu Thr Thr Trp Asp Asn Asp
 1 5 10 15

Cys Val Arg Gly Val Glu Met Gly Leu Pro Ala His Lys Leu His His
 20 25 30

Lys Ile Ile Arg Arg Lys Gly Val Ser Thr Gln Asn Asp Thr Glu Pro
 35 40 45

Pro Arg Thr Ser Arg Glu Asp Met
 50 55

<210> 163
 <211> 73
 <212> PRT
 <213> Homo sapien

<400> 163

Met Ala Trp Leu Gly Leu Arg Gly Leu Thr Phe Leu Pro Ser Tyr Ile
 1 5 10 15

Asn Lys Lys Asn Lys Thr Asn Ser Val Glu Val Leu Gly Trp Gln Asn
 20 25 30

104

Phe Trp Gly Val Ile Trp Arg Glu Asn Gly Arg Cys Phe Ser Gly Leu
 35 40 45

Leu Arg Ala Gly Leu Gly Ala Ala Trp Glu Pro Ser Thr Val Arg Val
 50 55 60

Thr Arg Ser Ser Ala Ser Val Met Val
 65 70

<210> 164
 <211> 72
 <212> PRT
 <213> Homo sapien

<400> 164

Asp Tyr Ala Glu Ser Pro Ala Ala Leu Ser Asn Gln Thr Ser Ala Val
 1 5 10 15

Val Pro Ile Leu Arg Pro Phe Ile Pro Val Phe Leu Leu Leu Phe
 20 25 30

His Leu Val Phe Gln Phe Ile Gln Asn Arg Ile Gln Ala Ile Thr Asn
 35 40 45

His Ser Ile Ala Gln Met Phe Leu Leu Thr Ser Pro Gln Ser His Pro
 50 55 60

Leu Pro Gln Asp Leu Pro Ser Ala
 65 70

<210> 165
 <211> 66
 <212> PRT
 <213> Homo sapien

<400> 165

Met Trp Met Leu Arg Phe Val His Leu Arg Trp Arg Leu Ser Ser Met
 1 5 10 15

Val Pro Pro Ser Ser Glu Leu Gln Pro Arg Met Glu Lys Ser Ser Ala
 20 25 30

Thr Val Gln Pro Tyr Gly Asn Ala Ser Ala His Ser Asp Leu Ala Ser
 35 40 45

105

Val Pro Ala Asn Arg Leu Val His Arg Ile Arg Pro Gly Ser Pro Gly
 50 55 60

Tyr Leu
 65

<210> 166
 <211> 126
 <212> PRT
 <213> Homo sapien

<400> 166

Met Leu Tyr Asp Gly Val Ile Ser Leu His His Ile Ala Leu Thr Val
 1 5 10 15

Thr Cys Ile Arg Ile His Gly Lys Ile Glu Asp Asn Leu Gln Gly Ala
 20 25 30

Ser Thr Ala Gln Leu Trp Cys Arg Arg Gly Asn His Ile Asn Gly Asp
 35 40 45

Ser Asn Asn Arg Ile Ile Pro Arg Arg Pro Pro Thr Leu Ile Tyr Thr
 50 55 60

Cys His Cys His Arg Ser Lys Ser Glu Asp Ser His Ile Gly Leu Gly
 65 70 75 80

His Ala Phe Gly Val Ala Pro Tyr Phe Arg Gly Tyr Ile Thr Met Met
 85 90 95

Trp Val Asp Thr Leu Lys Phe Ala Ala Leu Thr Arg His Leu Leu Leu
 100 105 110

Phe Arg Pro Leu Arg Arg Thr Phe Ile Lys Asn Pro Tyr Ile
 115 120 125

<210> 167
 <211> 69
 <212> PRT
 <213> Homo sapien

<400> 167

Met Ser Gln Glu Lys Asp Phe His Lys Val Met Ser Ser Leu Lys Ala
 1 5 10 15

106

Arg Thr Gly His Leu His Phe Phe Cys Gly Gly Arg Ser Ser Val Lys
 20 25 30

Val Gly Gln Ser Ile Phe Thr Ser Ser Val Ile Leu Gln Leu Leu Gln
 35 40 45

Ala Ile Trp Ala Tyr Thr Cys Lys Ser Gln Gly Met Arg Trp Leu Gly
 50 55 60

Leu Gly Ser Glu Ala
 65

<210> 168
 <211> 469
 <212> PRT
 <213> Homo sapien

<400> 168

Arg Ser Ser Lys Thr Ser Pro Asp Ile Ser His Gln Gln Ala Ala Ala
 1 5 10 15

Leu Leu His Thr Tyr Leu Lys Asn Leu Ser Pro Cys Ile Asn Ser Thr
 20 25 30

Pro Pro Ile Phe Gly Pro Leu Thr Thr Gln Thr Thr Ile Pro Val Ala
 35 40 45

Ala Pro Leu Cys Ile Ser Arg Gln Arg Pro Thr Gly Ile Pro Leu Gly
 50 55 60

Asn Leu Ser Pro Ser Arg Cys Ser Phe Thr Leu His Leu Arg Ser Pro
 65 70 75 80

Thr Thr His Ile Thr Glu Thr Ile Gly Ala Phe Gln Leu His Ile Thr
 85 90 95

Asp Lys Pro Ser Ile Asn Thr Asp Lys Leu Lys Asn Ile Ser Ser Asn
 100 105 110

Tyr Cys Leu Gly Arg His Leu Pro Ser Ile Ser Leu His Pro Trp Leu
 115 120 125

Pro Ser Pro Cys Ser Ser Asp Ser Pro Pro Arg Pro Ser Ser Arg Leu
 130 135 140

107

Leu Ile Pro Ser Pro Lys Asn Asn Ser Glu Arg Leu Leu Val Asp Thr
 145 150 155 160

Gln Arg Phe Leu Ile His His Glu Asn Arg Thr Ser Pro Ser Thr Gln
 165 170 175

Leu Pro His Gln Ser Pro Leu Gln Pro Leu Thr Ala Ala Ser Leu Ala
 180 185 190

Gly Ser Leu Gly Ile Trp Val Gln Asp Thr Pro Phe Ser Thr Pro His
 195 200 205

Leu Phe Thr Leu His Leu Gln Phe Cys Leu Thr Gln Gly Leu Phe Phe
 210 215 220

Leu Cys Gly Ser Ser Thr Tyr Met Cys Leu Pro Ala Asn Trp Thr Gly
 225 230 235 240

Thr Cys Thr Leu Val Phe Leu Thr Pro Lys Ile Gln Phe Ala Asn Gly
 245 250 255

Thr Glu Glu Leu Pro Val Pro Leu Met Thr Pro Thr Arg Gln Lys Arg
 260 265 270

Val Ile Pro Leu Ile Pro Leu Met Val Gly Leu Gly Leu Ser Ala Ser
 275 280 285

Thr Ile Ala Leu Gly Thr Gly Ile Ala Gly Ile Ser Thr Ser Val Thr
 290 295 300

Thr Phe Arg Ser Leu Ser Asn Asp Phe Ser Ala Ser Ile Thr Asp Ile
 305 310 315 320

Ser Gln Thr Leu Ser Val Leu Gln Ala Gln Val Asp Ser Leu Ala Ala
 325 330 335

Val Val Leu Gln Asn Arg Arg Gly Leu Asp Leu Leu Thr Ala Glu Lys
 340 345 350

Gly Gly Leu Cys Ile Phe Leu Asn Glu Glu Cys Cys Phe Tyr Leu Asn
 355 360 365

Gln Ser Gly Leu Val Tyr Asp Asn Ile Lys Lys Leu Lys Asp Arg Ala
 370 375 380

108

Gln Lys Leu Ala Asn Gln Ala Ser Asn Tyr Ala Glu Pro Pro Trp Ala
385 390 395 400

Leu Ser Asn Arg Met Ser Trp Val Leu Pro Ile Leu Ser Pro Leu Ile
405 410 415

Pro Ile Phe Leu Leu Leu Leu Phe Ala Pro Cys Ile Phe Cys Leu Val
420 425 430

Ser Gln Phe Ile Gln Asn Arg Ile Gln Ala Ile Thr Asn His Ser Ile
435 440 445

Ala Gln Met Phe Leu Leu Thr Thr Pro Gln Tyr His Pro Leu Pro Gln
450 455 460

Asp Leu Pro Ser Ala
465

<210> 169
<211> 243
<212> PRT
<213> Homo sapien

<400> 169

Met Thr Gly Arg Asn Asn Pro Ala Thr Ser Tyr Thr Asp Ala Gln His
1 5 10 15

Pro Gln Thr His Gln Lys His Thr Asn Gly Arg Thr Thr Arg Ala His
20 25 30

Lys Gln Ala Ala Gln Gln Ala Arg Ser Gln Gln Thr Arg Gln Ala Gln
35 40 45

Glu Gln Pro Thr Lys Glu Thr Asp Asn Thr Thr Glu Arg Arg Arg Gln
50 55 60

Arg Asn Ala Glu Gln Lys Asn Ala Gln Glu Ser Gln Gln Lys Gln Lys
65 70 75 80

His Pro Lys Gly Thr Glu Arg Lys Ala Glu Arg Asn Glu Thr Lys Glu
85 90 95

Glu Arg Arg Gln Glu Glu Lys Gln His Thr Asn Ala Asp Lys Glu Arg
100 105 110

109

Glu Arg Lys Thr Gln Thr Ser Arg Glu Thr Lys Thr Gly Asp Arg Gly
 115 120 125

Glu Glu Thr Arg Thr Ala Lys Arg Gln Lys Lys Glu Thr Lys Lys Gln
 130 135 140

Thr Thr Ala Arg Glu Asp Glu Lys Thr Asn Arg Arg Arg Arg Gln Glu
 145 150 155 160

Glu Thr Lys Thr Thr Lys Lys Arg Thr Ala Glu Asn Asn Ala Glu Arg
 165 170 175

Arg Lys Lys Lys Lys Arg Asp Gly Gln Gln Glu Thr Glu Arg Arg Asn
 180 185 190

Lys Asp Lys Arg Glu Glu Gln Asn Lys Arg Asp Lys Leu Arg Pro Thr
 195 200 205

Ser Glu Glu Arg His Lys Gln Glu Gln Gln Arg Ala Thr Gly Thr Arg
 210 215 220

Arg Ala Ala Ser Ser Gln Gly Asp Lys Arg Arg Glu Gln Arg His Asp
 225 230 235 240

Glu Lys Glu

<210> 170
 <211> 52
 <212> PRT
 <213> Homo sapien

<400> 170

Met Cys His Thr Pro Gly Ser Pro Arg Phe Phe Pro Leu Val Arg Leu
 1 5 10 15

Leu Pro Arg Cys Val Ile Phe Val Pro Cys Leu Phe Phe Leu Phe Ser
 20 25 30

Pro Phe Leu Ser Glu Cys Val Asn Gly Asn Glu Ser Ser Lys Asn Ser
 35 40 45

Ile Gly Gln Arg
 50

110

<210> 171
 <211> 167
 <212> PRT
 <213> Homo sapien
 <400> 171

Met Glu Val Thr Gly Val Ser Ala Pro Thr Val Thr Val Phe Ile Ser
 1 5 10 15

Ser Ser Leu Asn Thr Phe Arg Ser Glu Lys Arg Tyr Ser Arg Ser Leu
 20 25 30

Thr Ile Ala Glu Phe Lys Cys Lys Leu Glu Leu Leu Val Gly Ser Pro
 35 40 45

Ala Ser Cys Met Glu Leu Glu Leu Tyr Gly Val Asp Asp Lys Phe Tyr
 50 55 60

Ser Lys Leu Asp Gln Glu Asp Ala Leu Leu Gly Ser Tyr Pro Val Asp
 65 70 75 80

Asp Gly Cys Arg Ile His Val Ile Asp His Ser Gly Ala Arg Leu Gly
 85 90 95

Glu Tyr Glu Asp Val Ser Arg Val Glu Lys Tyr Thr Ile Ser Gln Glu
 100 105 110

Ala Tyr Asp Gln Arg Gln Asp Thr Val Arg Ser Phe Leu Lys Arg Ser
 115 120 125

Lys Leu Gly Arg Tyr Asn Glu Glu Glu Arg Ala Gln Gln Glu Ala Glu
 130 135 140

Ala Ala Gln Arg Leu Ala Glu Glu Lys Ala Gln Ala Ser Ser Ile Pro
 145 150 155 160

Val Gly Ser Arg Cys Glu Val
 165

<210> 172
 <211> 100
 <212> PRT
 <213> Homo sapien
 <400> 172

111

Met Cys Trp Ser Val Ser Ser Arg Gly Pro Arg Val Pro Ser Ala Pro
1 5 10 15

Thr Pro Ser Gly Pro Ala Leu Leu Pro Trp Asp Pro Thr Pro Pro Pro
20 25 30

Gly Asp Lys Lys Gly Gly Val Ala Pro Val Lys Lys Gly Gln Thr Pro
35 40 45

Pro Pro Asn Asn Ala Gly Pro Glu Lys Asn Asn Gln Arg Thr Ser Val
50 55 60

Phe Pro Leu Thr Cys Ser Lys Lys Asn Lys Lys Lys Lys Lys Lys Lys
65 70 75 80

Lys Lys Lys Lys Glu Pro Trp Gly Glu Asn His Gly Gly Thr Lys Gly
85 90 95

Leu Thr Arg Gly
100

<210> 173

<211> 358

<212> PRT

<213> Homo sapien

<400> 173

Met Pro Tyr His Ile Thr Ala Trp Ile Gln Gly Ser Thr Arg Arg Lys
1 5 10 15

His His Cys Gly Asp Thr Tyr Tyr Gly Thr Leu Gly Gly Ser Gln Glu
20 25 30

Thr Ser Lys Asn Asn Thr His Arg Ala Lys Lys Glu Gln Glu Asp Asn
35 40 45

Lys Lys Asp Gly Glu Ser Glu Gly Glu Tyr Thr His Lys Asp Gly Ala
50 55 60

Gln Gln Lys Glu Ala Glu Val His Thr Arg Gln Trp Val Tyr Thr Lys
65 70 75 80

Thr Gly Asp Arg Arg Ser Glu Ala Thr Gln Gln Arg Asn Gln His Thr
85 90 95

112

Asn Lys Lys Lys Pro His Pro Arg Arg Arg Arg Glu Arg Gly Gly Thr
 100 105 110

Gly Lys Thr Ala Gly Glu Gly Glu Glu Lys Lys Glu Arg Arg Gly Ala
 115 120 125

Ala Gln Lys Glu Arg Asp Glu Arg Thr Arg Arg Glu Arg Arg Glu Ala
 130 135 140

Glu Lys Lys Arg Asp Gln Gly Glu Arg Arg Ala Gln Ala Thr Gly Gln
 145 150 155 160

Ser Gly Arg Asn Thr Arg Gly Gly Glu Arg Glu Ser Glu Thr Arg Gln
 165 170 175

Arg Glu Lys Glu Gly Arg Glu Gly Arg Gly Gly His Gln Gly Glu Gly
 180 185 190

Lys Glu Lys Arg Gln Arg Lys Arg Arg Lys His Glu Glu Gly Arg Arg
 195 200 205

Arg Glu Arg Glu Gly His Glu Lys Thr Glu Arg Glu Arg Gly Glu Asp
 210 215 220

Thr Asp Ala Lys Arg Lys Arg Arg Arg Ser Lys Gly Arg Arg Lys Ser
 225 230 235 240

Lys Ser Arg Glu Arg Gln Thr Arg Glu Gly Thr Gln Asn His Arg Asp
 245 250 255

Asp Arg Lys Ser Arg Asn Glu Gly Lys Arg Ala Gly Thr Ala Arg Gln
 260 265 270

Arg His Lys Gln Asp Arg Lys Lys Arg Asn Glu Lys Arg Arg Asp Glu
 275 280 285

Ala Ala Thr Gln Arg Arg His Arg Gln Arg Glu Glu Arg Arg Asp Glu
 290 295 300

Gly Arg Arg Glu His Thr Lys Arg Gly Gln Arg Ser Lys Glu Arg Arg
 305 310 315 320

Asn Thr Arg Lys Arg Arg Asp Arg Arg Pro Arg Asp Thr Lys Gln Asp
 325 330 335

113

Asp Thr Glu Thr His Asp Gln Glu Arg Ala Gln Lys Glu Gln Thr Gln
 340 345 350

Lys Arg His Glu Glu Thr
 355

<210> 174
 <211> 48
 <212> PRT
 <213> Homo sapien

<400> 174

Met Leu Cys Gly Phe Glu Ala Ser Asn His Phe Thr His Ala Thr Gly
 1 5 10 15

Tyr Glu His Trp Val Thr Gln Ala Val Phe Leu Pro Leu Met Arg Glu
 20 25 30

Gly Ser Val Glu Ser Arg Leu Cys Val Val Pro Gly His Phe His Pro
 35 40 45

<210> 175
 <211> 64
 <212> PRT
 <213> Homo sapien

<400> 175

Met Arg Asp Gly Lys Thr Gln Leu Ala Ala Arg His Gly Arg Thr His
 1 5 10 15

Val Arg Arg Thr His Arg His Ala Pro Leu Pro Trp Asn Ser Arg Thr
 20 25 30

Ala Tyr Pro Ser Cys His Leu Pro Ser Gln Gln Arg Phe Asn Arg Arg
 35 40 45

Thr Ile Asp Ala Gly Gly Met Gln Gly Asn Val Phe Leu Met Leu Pro
 50 55 60

<210> 176
 <211> 64
 <212> PRT
 <213> Homo sapien

<400> 176

Met Arg Ser Trp Thr Lys Asp Ile Tyr Ser Phe Ile Gln Tyr Ser Cys

114

1 5 10 15

Val Cys Val Leu Glu Thr Gly Phe Glu Arg Arg Met Val Arg Ser Ala
 20 25 30

Val Arg Ser Arg Gly Arg Ser Gly Trp Thr Lys Thr Arg Thr Pro Asp
 35 40 45

Leu Gly Ser Pro Ala Asp Leu Asp Arg Ile Leu Pro Asp Ser Val Pro
 50 55 60

<210> 177
<211> 254
<212> PRT
<213> Homo sapien

<400> 177

Met Arg Arg Glu Arg Pro Arg Asp Arg Arg Arg Arg Lys Gly Ala Gly
1 5 10 15

Thr Arg Glu Arg Glu Arg Glu Gly Glu Glu Glu Ala Gln Ala Ala Pro
 20 25 30

Gly Gln Ala Glu Ser Gly Arg Gly Gly Lys Ser Asp Arg Glu Ala Arg
 35 40 45

Ala His Lys Lys Ala Ser Asp Thr Arg Arg Asp Thr Ala Gln Ser Lys
 50 55 60

Gly Glu Asp Pro Glu Arg Glu Gly Arg Arg Arg Glu Gly Asp Ala Asp
65 70 75 80

Pro Arg Thr Arg Gly Gln Ala Arg Arg His Gly Arg Gln Arg Glu Glu
 85 90 95

Gly Gly Gly Lys Lys Arg Glu Arg Thr Arg Gly Pro Arg Ser Gly Asp
 100 105 110

Glu Glu Arg Arg Ala Gln Ala Gln Arg Gln Ala Arg Arg Arg Gly Gly
 115 120 125

Glu Arg Ala Ser Gly Arg Ala Arg Arg Ser Val Thr Lys Gly Gly Arg
 130 135 140

Ser Arg Pro His Ser Arg Arg Ala Arg Arg Gln Glu Thr Glu Glu Arg

115

145

150

155

160

Glu Arg Gly Asp Arg Gly Arg Thr Arg Glu Lys Gly Arg Thr Arg Arg
 165 170 175

Arg Asn Arg Ala Arg Gly Arg Arg Pro Pro Ser Thr Arg Gln Ala Gly
 180 185 190

Thr Thr Arg Gln Thr Glu Arg Lys Gln Ala Arg Ala Arg Pro Asn Arg
 195 200 205

Pro Thr Ser Glu Ala Arg Ala Lys Arg Gln Ala His Gln Asp Ala Asn
 210 215 220

Gln Ala Ala Asp Glu Glu Gly Gln Asp His Lys Arg Arg Arg Pro Gly
 225 230 235 240

Arg Glu Glu Arg Gly Gln Pro Glu Ala Ala His Thr Asn Ser
 245 250

<210> 178
 <211> 58
 <212> PRT
 <213> Homo sapien

<400> 178

Met Val Cys Val Met Leu Pro Gln Pro His Ala Ser Arg Gly Cys Cys
 1 5 10 15

Cys Ala Gln Asp Val Cys Gln Gly Ala Pro Glu Gly Val Leu Arg Pro
 20 25 30

Leu Leu Thr Ile Gly Ala Arg Leu Glu Thr Ser Arg Gly Ala Leu Thr
 35 40 45

Gly Pro Val Asn Gly Lys Arg Ser Leu Arg
 50 55

<210> 179
 <211> 170
 <212> PRT
 <213> Homo sapien

<400> 179

Met Ser Glu Leu Thr His Ile Asn Cys Val Ala Leu Thr Ala Arg Phe
 1 5 10 15

116

Pro Val Gly Lys Pro Val Val Pro Ala Ala Leu Met Asn Arg Pro Thr
 20 25 30

Arg Gly Glu Arg Arg Phe Ala Tyr Trp Ala Leu Phe Arg Phe Leu Ala
 35 40 45

His Ala Leu Ala Ala Leu Gly Arg Ser Ala Ala Ala Ser Gly Ile Ser
 50 55 60

Ser Leu Lys Gly Gly Asn Thr Val Ile His Arg Ile Arg Gly Ala Arg
 65 70 75 80

Arg Arg Glu His Val Ser Lys Arg Pro Ala Lys Gly Gln Glu Pro Ala
 85 90 95

Lys Gly Arg Val Ala Gly Val Phe Pro His Asn Ile Arg Ala Tyr Glu
 100 105 110

His Leu Pro Lys Ser Arg Pro Pro Arg Ala Arg Thr Ser Pro His Ser
 115 120 125

Leu Asn His Arg Asn Asp Leu Phe Pro Phe Thr Gly Pro Val Arg Ala
 130 135 140

Pro Arg Asp Val Ser Asn Leu Ala Pro Ile Val Asn Ser Gly Arg Ser
 145 150 155 160

Thr Pro Ser Gly Ala Pro Ala His Thr Ser
 165 170

<210> 180
 <211> 111
 <212> PRT
 <213> Homo sapien

<400> 180

Met Thr Leu Asn Glu His Ala Ala Phe Lys His Leu Phe Asn Lys Ala
 1 5 10 15

His Leu Ala Leu Pro Leu Ile His Leu Thr Leu Ser Gly His Arg Thr
 20 25 30

Cys Phe Arg Glu His Arg Val Gly Gly Lys Val Thr Asp Gln Gln Asp
 35 40 45

117

Pro Lys Ala Glu Glu Phe Phe Leu Val Ala Asn Lys Met Lys Ser Leu
 50 55 60

Pro Cys Leu Leu Leu Ser Thr Gln Thr Arg Gln Pro Ser Asp Phe Ser
 65 70 75 80

Ile Phe Ser Pro Pro Phe Pro Pro Phe Tyr Ser Thr Lys Pro Pro Leu
 85 90 95

Ser Ser Trp Pro Val Leu Asn Glu Leu Leu Gly Thr Cys Pro Arg
 100 105 110

<210> 181

<211> 77

<212> PRT

<213> Homo sapien

<400> 181

Met Gly Gly Asn Gln Phe Gln Pro Glu Pro Phe Gly Gln Val Thr Pro
 1 5 10 15

Ala Phe Phe Phe Phe Phe Leu Gly Met Glu Ser Arg Cys Ile Pro Arg
 20 25 30

Leu Glu Cys Ser Gly Ala Ile Ser Ala His Cys Lys Leu His Leu Pro
 35 40 45

Gly Phe Thr Pro Phe Ser Cys Leu Arg Leu Pro Ser Ser Trp Asp Tyr
 50 55 60

Arg Arg Pro Pro Pro His Arg Ala Asn Phe Leu Tyr Phe
 65 70 75

<210> 182

<211> 75

<212> PRT

<213> Homo sapien

<400> 182

Arg Pro Ser Ala Val Ala His Ala Cys Asn Pro Ser Thr Leu Gly Gly
 1 5 10 15

Gln Gly Gly Trp Ile Thr Arg Ser Gly Asp Gln Asp His Pro Gly Ala
 20 25 30

118

His Gly Glu Thr Pro Ser Leu Leu Lys Ile Gln Lys Ile Ser Pro Val
 35 40 45

Trp Trp Trp Ala Pro Val Val Pro Ala Thr Arg Glu Ala Glu Ala Gly
 50 55 60

Glu Trp Arg Glu Pro Gly Arg Trp Ser Leu Gln
 65 70 75

<210> 183
 <211> 147
 <212> PRT
 <213> Homo sapien

<400> 183

Met Lys Tyr Val Leu Val Tyr Phe His Ala Ala Asp Lys Asp Ile Pro
 1 5 10 15

Glu Thr Gly Glu Lys Lys Arg Phe Ser Trp Thr Tyr Ser Ser Thr Trp
 20 25 30

Leu Gly Arg Pro Gln Asn His Gly Glu Arg Arg Lys Ala Leu Leu Thr
 35 40 45

Trp Trp Gln Gln Glu Lys Thr Arg Lys Lys Gln Lys Arg Lys Ser Leu
 50 55 60

Ile Ile Pro Ser Asp Leu Met Arg Arg Ile His Tyr Tyr Lys Asn Gly
 65 70 75 80

Met Val Lys Thr Ser Pro His Asp Ser Ile Thr Ser Pro Gly Ser Leu
 85 90 95

Pro Gln Arg Val Gly Ile Leu Gly Asp Thr Ile Gln Val Glu Ile Trp
 100 105 110

Val Gly Thr Gln Pro Asn His Ile Ile Leu Pro Leu Ala Pro Ser Lys
 115 120 125

Ser His Val Leu Thr Phe Gln Asn Gln Ser Cys Leu His Asn Ser Pro
 130 135 140

Pro Lys Ser
 145

119

<210> 184
 <211> 94
 <212> PRT
 <213> Homo sapien

<400> 184

Trp Leu Lys Arg Ala Asn Ile Glu Leu Arg Leu Trp Leu Gln Arg Val
 1 5 10 15

Glu Ala Pro Ser Leu Gly Ser Phe His Met Val Leu Ser Leu Gln Val
 20 25 30

His Arg Ser Gln Glu Leu Arg Phe Gly Asn Leu His Leu Asp Phe Arg
 35 40 45

Arg Cys Met Glu Met Pro Gly Cys Pro Gly Lys Ser Trp His Gln Gly
 50 55 60

His Ser Pro Tyr Gly Lys Leu Leu Pro Gly His Cys Gly Ser Lys Leu
 65 70 75 80

Trp Gly Gln Ser Pro Thr Gln Ser Pro Ala Trp Gly Thr Ala
 85 90

<210> 185
 <211> 17
 <212> PRT
 <213> Homo sapien

<400> 185

Met Leu Ser Ser Gly Ala Cys Asp Gly Ser Ala Pro Leu Gln Pro Cys
 1 5 10 15

Ala

<210> 186
 <211> 125
 <212> PRT
 <213> Homo sapien

<400> 186

Met Ser Pro Leu Lys Asn Pro Gln Pro Pro Phe Phe Phe Phe Phe Phe
 1 5 10 15

Phe Phe Phe Glu Pro Gly Val Ser Ile Leu Thr Ser Val Ala Pro Lys

120

20

25

30

Val Lys Cys Thr Val Ala Pro Ile Thr Gly Leu Thr Ala Ser Pro Gly
 35 40 45

Pro Pro Gly Leu Thr Val Asn Pro Phe Cys Leu Ser Leu Pro Ser Arg
 50 55 60

Val Ala Gly Thr Trp Asp Tyr Arg Gln Ala His His Thr Pro Thr Thr
 65 70 75 80

Phe Val Phe Phe Phe Phe Leu Val Glu Ile Gly Val Pro Pro Cys Tyr
 85 90 95

Pro Gly Trp Ser Arg Thr Pro Val Val Lys Gln Ser Ser Ile Thr Leu
 100 105 110

Arg Arg Ser Ser Met His Leu Glu Thr His Ser Pro Ile
 115 120 125

<210> 187
 <211> 84
 <212> PRT
 <213> Homo sapien

<400> 187

Met His Ser Gly Trp Glu Trp Trp Leu Met Pro Val Ile Pro Ala Val
 1 5 10 15

Trp Glu Ala Glu Val Gly Arg Leu Phe Asp His Arg Ser Ser Arg Pro
 20 25 30

Ala Gly Val Thr Trp Gln Asp Pro Asn Leu Tyr Gln Lys Lys Lys Lys
 35 40 45

Tyr Lys Ser Cys Arg Gly Val Val Cys Leu Pro Val Val Pro Ser Pro
 50 55 60

Ser Tyr Ser Thr Trp Glu Ala Glu Ala Glu Gly Ile His Arg Glu Pro
 65 70 75 80

Arg Arg Ala Arg

<210> 188

121

<211> 89
 <212> PRT
 <213> Homo sapien

<400> 188

Met Cys Phe Val Lys Gln Met Leu Glu Gly Ser Met Leu Val Lys Ser
 1 5 10 15

His His Gln Ser Leu Ile Ser Ser Asn Gln Gly His Lys His Cys Gly
 20 25 30

Arg Pro Gln Gly Pro Leu Pro Arg Lys Thr Arg Asp Leu Cys Ser Leu
 35 40 45

Val Tyr Leu Leu Thr Phe Pro Pro Leu Leu Ser His Asp Pro Ala Lys
 50 55 60

Tyr Pro Ser Val Arg Asn Thr Gln Glu Leu Ser Lys Lys Lys Lys Glu
 65 70 75 80

Glu Lys Lys Lys Lys Lys Gly Gly Gly
 85

<210> 189
 <211> 917
 <212> PRT
 <213> Homo sapien

<400> 189

Ala Ala Leu Ser Lys Cys Lys Arg Thr Glu Ile Ile Thr Asn Tyr Leu
 1 5 10 15

Ser Asp His Ser Ala Ile Lys Leu Glu Leu Arg Ile Lys Asn Leu Thr
 20 25 30

Gln Ser Arg Ser Thr Thr Trp Lys Leu Asn Asn Leu Leu Leu Asn Asp
 35 40 45

Tyr Trp Val His Asn Glu Met Lys Ala Glu Ile Lys Met Phe Phe Glu
 50 55 60

Thr Asn Glu Asn Lys Asp Thr Thr Tyr Gln Asn Leu Trp Asp Ala Phe
 65 70 75 80

Lys Ala Val Cys Arg Gly Lys Phe Ile Ala Leu Asn Ala His Lys Arg
 85 90 95

122

Lys Gln Glu Arg Ser Lys Ile Asp Thr Leu Thr Ser Gln Leu Lys Glu
 100 105 110

Leu Glu Lys Gln Glu Gln Thr His Ser Lys Ala Ser Arg Arg Gln Glu
 115 120 125

Ile Thr Lys Ile Arg Ala Glu Leu Lys Glu Ile Glu Thr Gln Lys Thr
 130 135 140

Leu Gln Lys Ile Asn Glu Ser Arg Ser Trp Phe Phe Glu Arg Ile Asn
 145 150 155 160

Lys Ile Asp Arg Pro Leu Ala Arg Leu Ile Lys Lys Lys Arg Glu Lys
 165 170 175

Asn Gln Ile Asp Thr Ile Lys Asn Asp Lys Gly Asp Ile Thr Thr Asp
 180 185 190

Pro Thr Glu Ile Gln Thr Thr Ile Arg Glu Tyr Tyr Lys His Leu Tyr
 195 200 205

Ala Asn Lys Leu Glu Asn Leu Glu Glu Met Asp Lys Phe Leu Asp Thr
 210 215 220

Asp Thr Leu Pro Arg Leu Asn Gln Glu Glu Val Glu Ser Leu Asn Arg
 225 230 235 240

Pro Ile Thr Gly Ala Glu Ile Val Ala Ile Ile Asn Ser Leu Pro Thr
 245 250 255

Lys Lys Ser Pro Gly Pro Asp Gly Phe Thr Ala Glu Phe Tyr Gln Arg
 260 265 270

Tyr Lys Glu Glu Leu Val Pro Phe Leu Leu Lys Leu Phe Gln Ser Ile
 275 280 285

Glu Lys Glu Gly Ile Leu Pro Asn Ser Phe Tyr Glu Ala Ser Ile Ile
 290 295 300

Leu Ile Pro Lys Pro Gly Arg Asp Thr Thr Lys Lys Glu Asn Phe Arg
 305 310 315 320

Pro Ile Ser Leu Met Asn Ile Asp Ala Lys Ile Leu Asn Lys Ile Leu

123

325

330

335

Ala Lys Arg Ile Gln Gln His Ile Lys Lys Leu Ile His His Asp Gln
 340 345 350

Val Gly Phe Ile Pro Gly Met Gln Gly Trp Phe Asn Ile Arg Lys Ser
 355 360 365

Ile Asn Val Ile Gln His Ile Asn Arg Ala Lys Asp Lys Asn His Met
 370 375 380

Ile Ile Ser Ile Asp Ala Glu Lys Ala Phe Asp Lys Ile Gln Gln Pro
 385 390 395 400

Phe Met Leu Lys Thr Leu Asn Lys Leu Gly Ile Asp Gly Thr Tyr Phe
 405 410 415

Lys Ile Ile Arg Ala Ile Tyr Asp Lys Pro Thr Ala Asn Ile Ile Leu
 420 425 430

Asn Gly Gln Lys Leu Glu Ala Phe Pro Leu Lys Thr Gly Thr Arg Gln
 435 440 445

Gly Cys Pro Leu Ser Pro Leu Leu Phe Asn Ile Val Leu Glu Val Leu
 450 455 460

Ala Arg Ala Ile Arg Gln Glu Lys Glu Ile Lys Gly Ile Gln Leu Gly
 465 470 475 480

Lys Glu Glu Val Lys Leu Ser Leu Phe Ala Asp Asp Met Ile Val Tyr
 485 490 495

Leu Glu Asn Pro Ile Val Ser Ala Gln Asn Leu Leu Lys Leu Ile Ser
 500 505 510

Asn Phe Ser Lys Val Ser Gly Tyr Lys Ile Asn Val Gln Lys Ser Gln
 515 520 525

Ala Phe Leu Tyr Thr Asn Asn Arg Gln Thr Glu Ser Gln Ile Met Ser
 530 535 540

Glu Leu Pro Phe Thr Ile Ala Ser Lys Arg Ile Lys Tyr Leu Gly Ile
 545 550 555 560

124

Gln Leu Thr Arg Asp Val Lys Asp Leu Phe Lys Glu Asn Tyr Lys Pro
 565 570 575

Leu Leu Lys Glu Ile Lys Glu Asp Thr Asn Lys Trp Lys Asn Ile Pro
 580 585 590

Cys Ser Trp Val Gly Arg Ile Asn Ile Val Lys Met Ala Ile Leu Pro
 595 600 605

Lys Val Ile Tyr Arg Phe Asn Ala Ile Pro Ile Lys Leu Pro Met Thr
 610 615 620

Phe Phe Thr Glu Leu Glu Lys Thr Thr Leu Lys Phe Ile Trp Asn Gln
 625 630 635 640

Lys Arg Ala Arg Ile Ala Lys Ser Ile Leu Ser Gln Lys Asn Lys Ala
 645 650 655

Gly Gly Ile Thr Leu Pro Asp Phe Lys Leu Tyr Tyr Lys Ala Thr Val
 660 665 670

Thr Lys Thr Ala Trp Tyr Trp Tyr Gln Asn Arg Asp Ile Asp Gln Trp
 675 680 685

Asn Arg Thr Glu Pro Ser Glu Ile Met Pro His Ile Tyr Asn Tyr Leu
 690 695 700

Ile Phe Asp Lys Pro Glu Lys Asn Lys Gln Trp Gly Lys Asp Ser Leu
 705 710 715 720

Phe Asn Lys Trp Cys Trp Glu Asn Trp Leu Ala Ile Cys Arg Lys Leu
 725 730 735

Lys Leu Asp Pro Phe Leu Thr Pro Tyr Thr Lys Ile Asn Ser Arg Trp
 740 745 750

Ile Lys Asp Leu Asn Val Arg Pro Lys Thr Ile Lys Thr Leu Glu Glu
 755 760 765

Asn Leu Gly Ile Thr Ile Gln Asp Ile Gly Val Gly Lys Asp Phe Met
 770 775 780

Ser Lys Thr Pro Lys Ala Met Ala Thr Lys Ala Lys Ile Asp Lys Trp
 785 790 795 800

125

Asp Leu Ile Lys Leu Lys Ser Phe Cys Thr Ala Lys Glu Thr Thr Ile
805 810 815

Arg Val Asn Arg Gln Pro Thr Thr Trp Glu Lys Ile Phe Ala Thr Tyr
820 825 830

Ser Ser Asp Lys Gly Leu Ile Ser Arg Ile Tyr Asn Glu Leu Lys Gln
835 840 845

Ile Tyr Lys Lys Lys Thr Asn Asn Pro Ile Lys Lys Trp Ala Lys Asp
850 855 860

Met Asn Arg His Phe Ser Lys Glu Asp Ile Tyr Ala Ala Lys Lys His
865 870 875 880

Met Lys Lys Cys Ser Ser Ser Leu Ala Ile Arg Glu Met Gln Ile Lys
885 890 895

Thr Thr Met Arg Tyr His Leu Thr Pro Val Arg Met Ala Ile Ile Lys
900 905 910

Lys Ser Gly Asn Asn
915

<210> 190

<211> 110

<212> PRT

<213> Homo sapien

<400> 190

Met Lys Cys Cys Val Glu Asn Cys Glu Arg Asn Asn Thr Phe His Thr
1 5 10 15

Thr Gly Thr Arg Tyr Glu Pro Leu Ser Tyr Ala Gln Pro Phe Phe Phe
20 25 30

Phe Ser Phe Phe Phe Phe Leu Leu Ser Phe Leu Ser Phe Phe Leu
35 40 45

Ser Phe Leu Leu Phe Leu Ser Leu Ser Leu Ser Leu Ser Phe Phe Leu
50 55 60

Pro Phe Phe Leu Ser Phe Ser Gln Ser Val Thr Pro Gly Trp Ser Ala
65 70 75 80

126

Val Ala Leu Ser Gln Leu Thr Ala Thr Ser Asp Ser Ser Val Gln Ala
85 90 95

Ile Leu Leu Pro Leu Pro Pro Lys Val Leu Arg Leu Gln Ala
100 105 110

<210> 191
<211> 43
<212> PRT
<213> Homo sapien

<400> 191

Met Gly Ala Thr Thr Gly Gly Gly Gly Arg Gly Arg Gln Gly Glu Glu
1 5 10 15

Ala Glu Ala Gly Glu Lys Arg Gly Glu Gln Gly Ala Val Trp Arg Gly
20 25 30

Lys Glu Arg Glu Arg Gly Ala Arg Ala Arg Arg
35 40

<210> 192
<211> 61
<212> PRT
<213> Homo sapien

<400> 192

Met Ala Lys Glu Leu Pro Gln Ala Leu Phe Phe Val Phe Phe Phe Phe
1 5 10 15

Leu Phe Val Leu Arg Trp Asn Leu His Phe Met Ser Pro Pro Gly Trp
20 25 30

Ser Ala Val Ala Ala Asp Leu Arg Leu Thr Ala Thr Phe Thr Cys Gln
35 40 45

Gly Ser Ser Asp Ser Pro Ala Ser Val Ser Gln Asn Ser
50 55 60

<210> 193
<211> 57
<212> PRT
<213> Homo sapien

<400> 193

Met Leu Phe Ala Trp Leu Ile Ser Pro Gly Thr Pro Ser Ile Arg Tyr

127

1 5 10 15

Glu Ile Ala Cys Met Leu Leu His Lys Val Thr Asp Arg Trp Gln Arg
20 25 30

Ser Thr Asn Ala Ala Pro Gly Arg Thr Thr His Cys Asp Lys Gln Asp
35 40 45

Leu Pro Gly Arg Pro Ile Leu Ser Thr
50 55

<210> 194
<211> 61
<212> PRT
<213> Homo sapien

<400> 194

Met Pro Leu His Ser Ser Leu Gly Asn Ile Val Arg Ser Cys Leu Lys
1 5 10 15

Asn Asn Asn Asn Lys Ile Gly Arg Ala Arg Trp Leu Thr Pro Val Ile
20 25 30

Pro Ala Leu Trp Glu Ala Lys Ala Gly Gly Ser Arg Gly Gln Glu Ile
35 40 45

Lys Thr Ile Leu Ala Asn Thr Val Lys Pro His Leu Tyr
50 55 60

<210> 195
<211> 75
<212> PRT
<213> Homo sapien

<400> 195

Arg Pro Ser Ala Val Ala His Ala Cys Asn Pro Ser Thr Leu Gly Gly
1 5 10 15

Gln Gly Gly Trp Ile Thr Arg Ser Gly Asp Gln Asp His Pro Gly Ala
20 25 30

His Gly Glu Thr Pro Ser Leu Leu Lys Ile Gln Lys Ile Ser Pro Val
35 40 45

Trp Trp Trp Ala Pro Val Val Pro Ala Thr Arg Glu Ala Glu Ala Gly
50 55 60

128

Glu Trp Arg Glu Pro Gly Arg Trp Ser Leu Gln
65 70 75

<210> 196
<211> 69
<212> PRT
<213> Homo sapien

<400> 196

Met Ser His His Ala Arg Pro His Leu Phe Phe Ile Arg Ser Ser Val
1 5 10 15

Gly Arg His Leu His Cys Phe Gln Ile Leu Ala Ile Val Asn Ser Ala
20 25 30

Ala Ile Asn Ile Arg Val Gln Thr Ser Leu Pro His Leu Ile Ser Phe
35 40 45

Leu Leu Gly Ile Tyr Leu Ala Val Glu Leu Leu Asp His Met Val Ala
50 55 60

Leu Phe Leu Val Phe
65

<210> 197
<211> 157
<212> PRT
<213> Homo sapien

<400> 197

Met Val Cys Glu Gln Thr Leu Gly Ser Val Val Val Trp Asn Met Trp
1 5 10 15

Ser Gly Lys Thr Asn Ile His His Gln Gly Thr Ser Phe Arg Thr Gln
20 25 30

Asp Leu Pro Pro Arg Leu Phe Phe Leu Phe Phe Phe Ser Glu Gln Ser
35 40 45

Leu Leu Cys Tyr Ile Gly Ala Gly Val Arg Cys His Asp Leu Ser Ser
50 55 60

Leu Gln Cys Leu Pro Ser Arg Phe Lys Gln Phe Leu Cys Leu Ser Leu
65 70 75 80

129

Pro Ser Ser Trp Asp Thr Gly Ala Arg His His Thr Gln Leu Ile Phe
 85 90 95

Ala Val Leu Val Glu Thr Gly Phe Cys His Val Gly Gln Ala Gly Leu
 100 105 110

Glu Leu Leu Ala Ser Ser Asp Leu Pro Ile Leu Ala Ser Gln Ser Ala
 115 120 125

Gly Thr Thr Gly Val Ser His Arg Thr Gln Leu Phe Phe Val Tyr Phe
 130 135 140

His Leu Leu Leu Cys Pro His His Phe Ser Leu Ser Leu
 145 150 155

<210> 198
 <211> 101
 <212> PRT
 <213> Homo sapien

<400> 198

Phe Phe Ser Glu Gln Ser Leu Leu Cys Tyr Ile Gly Ala Gly Val Arg
 1 5 10 15

Cys His Asp Leu Ser Ser Leu Gln Cys Leu Pro Ser Arg Phe Lys Gln
 20 25 30

Phe Leu Cys Leu Ser Leu Pro Ser Ser Trp Asp Tyr Arg Cys Thr Pro
 35 40 45

Pro His Pro Ala Asn Phe Ala Val Leu Val Glu Thr Gly Phe Cys His
 50 55 60

Val Gly Gln Ala Gly Leu Glu Leu Leu Ala Ser Ser Asp Leu Pro Ile
 65 70 75 80

Leu Ala Ser Gln Ser Ala Gly Thr Thr Gly Val Ser His Arg Thr Gln
 85 90 95

Leu Phe Phe Val Tyr
 100

<210> 199
 <211> 79
 <212> PRT

130

<213> Homo sapien

<400> 199

Met Ser Phe Leu Phe Leu Ser Cys Phe Phe Phe Ser Phe Ser Phe Ser
 1 5 10 15

Thr Phe Leu Phe Ser Phe Phe Ile Ser Cys Arg Phe Phe Cys Phe Leu
 20 25 30

Leu Cys Phe Leu Phe Leu Phe Cys Leu Ala Leu Ala Phe Asp Phe Leu
 35 40 45

Phe Thr Leu Phe Val Leu Leu Cys Leu Ser Ser Phe Val Phe Cys Leu
 50 55 60

Ser Leu Leu Phe Phe Ser Leu Val Leu Trp Val Cys Leu Leu Ser
 65 70 75

<210> 200

<211> 113

<212> PRT

<213> Homo sapien

<400> 200

Met Thr Leu Asn Glu His Ala Ala Phe Lys His Leu Phe Asn Lys Ala
 1 5 10 15

His Leu Ala Leu Pro Leu Ile His Leu Thr Leu Ser Gly His Arg Thr
 20 25 30

Cys Phe Arg Glu His Arg Val Gly Gly Lys Val Thr Asp Gln Gln Asp
 35 40 45

Pro Lys Ala Glu Glu Phe Phe Leu Val Ala Asn Arg Met Lys Ser Leu
 50 55 60

Pro Cys Leu Leu Leu Ser Thr Gln Thr Arg Gln Pro Ser Asp Phe Ser
 65 70 75 80

Ile Phe Ser Pro Pro Phe Pro Pro Phe Tyr Ser Thr Lys Pro Pro Leu
 85 90 95

Ser Ser Trp Pro Val Leu Asn Glu Leu Leu Gly Thr Cys Pro Gly Gly
 100 105 110

131

Arg

<210> 201
 <211> 108
 <212> PRT
 <213> Homo sapien

<400> 201

Met Ile Arg Thr Met Ile Ser Ser Gly Glu Glu Val Cys Gln Tyr Leu
 1 5 10 15

Met Arg Cys Asn Arg Leu Gly Thr Ala Gly Ala Asn Ser Ala Val Gln
 20 25 30

Asp Arg Trp Ser Ala Ile Ser Pro Ile Thr Ser Ser Cys Ser Cys His
 35 40 45

Thr Ser Gln Glu Lys Lys Lys Glu Lys Lys Met Gly Val Gly Gly Ile
 50 55 60

His Tyr Val Gly Ala Asn Arg Arg Ala Thr Pro Gly Gly Val Arg Met
 65 70 75 80

Trp Gly Val Cys Ala Ala Thr Thr Ile Cys Pro Pro Pro Ser Ile Ser
 85 90 95

Gly Ala Glu Thr Gly Gln Lys Gly Arg Glu Ala Thr
 100 105

<210> 202
 <211> 51
 <212> PRT
 <213> Homo sapien

<400> 202

Met Ser Tyr Arg Pro Ala Phe Ser Ala Trp Ala Trp Trp Phe Tyr Arg
 1 5 10 15

Pro Ile Ile Leu Ala Leu Trp Glu Ala Pro Gly Gly Trp Ile Thr Arg
 20 25 30

Gly Gln Gly Phe Lys Thr Pro Pro Gly Pro Asp Gly Glu Asn Pro His
 35 40 45

Leu Leu Pro

132

50

<210> 203
 <211> 117
 <212> PRT
 <213> Homo sapien

<400> 203

Phe Cys Gly Met Asn Ile Ala Asn Leu Ser Ala Gln Phe Pro Phe Phe
 1 5 10 15

Phe Phe Phe Leu Gly Gln Ser Leu Ala Leu Ser Leu Arg Leu Glu Cys
 20 25 30

Ile Gly Ala Val Ser Thr His Cys Asn Leu Arg Leu Trp Asp Ser Ser
 35 40 45

Asn Ser Pro Ala Ser Ala Ser Gln Ile Ala Gly Thr Thr Gly Met His
 50 55 60

Tyr His Ala Gln Ile Ile Phe Val Phe Leu Val Glu Thr Gly Val Ser
 65 70 75 80

Glu Thr Gly Phe His His Val Gly Gln Ala Asp Leu Glu Leu Leu Thr
 85 90 95

Ser Gly Asp Pro Pro Thr Leu Ala Ser Gln Ser Ala Ser Ile Met Gly
 100 105 110

Val Asn His His His
 115

<210> 204
 <211> 223
 <212> PRT
 <213> Homo sapien

<400> 204

Glu Arg Gly Leu Pro Gly Val Ala Gly Ala Val Gly Glu Pro Gly Pro
 1 5 10 15

Leu Gly Ile Ala Gly Pro Pro Gly Ala Arg Gly Pro Pro Gly Ala Val
 20 25 30

Gly Ser Pro Gly Val Asn Gly Ala Pro Gly Glu Ala Gly Arg Asp Gly
 35 40 45

133

Asn Pro Gly Asn Asp Gly Pro Pro Gly Arg Asp Gly Gln Pro Gly His
50 55 60

Lys Gly Glu Arg Gly Tyr Pro Gly Asn Ile Gly Pro Val Gly Ala Ala
65 70 75 80

Gly Ala Pro Gly Pro His Gly Pro Val Gly Pro Ala Gly Lys His Gly
85 90 95

Asn Arg Gly Glu Thr Gly Pro Ser Gly Pro Val Gly Pro Ala Gly Ala
100 105 110

Val Gly Pro Arg Gly Pro Ser Gly Pro Gln Gly Ile Arg Gly Asp Lys
115 120 125

Gly Glu Pro Gly Glu Lys Gly Pro Arg Gly Leu Pro Gly Leu Lys Gly
130 135 140

His Asn Gly Leu Gln Gly Leu Pro Gly Ile Ala Gly His His Gly Asp
145 150 155 160

Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly Pro Arg Gly Pro Ala
165 170 175

Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly His Pro Gly
180 185 190

Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly Pro
195 200 205

Ala Gly Pro Pro Gly Pro Pro Gly Pro Ser Trp Gly Pro Pro Gly
210 215 220

<210> 205

<211> 59

<212> PRT

<213> Homo sapien

<400> 205

Met Leu Lys Val Gly Ala Glu His Ile His Phe Leu Phe Val Ile Leu
1 5 10 15

Gln Val Thr Phe Arg Pro Ser Gly His Ile Pro Cys Asn Val Lys Glu
20 25 30

Val Val Ile Thr Asn Gln Lys Asn Lys Thr Lys
50 55

```
<210> 206
<211> 53
<212> PRT
<213> Homo sapien
```

<400> 206

Met Lys Phe Gly Ala Met Thr Arg Ile Gly Val Pro Pro Leu Gly Asp
1 5 10 15

Gln Ser Pro Ser Ser Cys Ser Leu Leu Arg Glu Lys Asp Leu Pro Arg
20 25 30

Thr Ser Gly Pro Gln Thr Asp Gln Pro Lys Glu His Leu Thr Asn Phe
35 40 45

Lys Ser Gly Thr Arg
50

<210>	207
<211>	135
<212>	PRT
<213>	Homo sapien

<400> 207

Met Cys Phe Val Lys Gln Met Leu Glu Gly Ser Met Leu Val Lys Ser

1 5 10 15

His His Gln Ser Leu Ile Ser Ser Asn Gln Gly His Lys His Cys Gly
20 25 30

Arg Pro Gln Gly Pro Leu Pro Arg Lys Thr Arg Asp Leu Cys Ser Leu
35 40 45

Val Tyr Leu Leu Thr Phe Pro Pro Leu Leu Ser His Asp Pro Ala Lys
50 55 60

Tyr Pro Ser Val Arg Asn Thr Gln Glu Leu Ser Lys Lys Lys Asn His
65 70 75 80

135

Lys Pro Lys Lys Lys Arg Leu Gly Asp Pro Trp Ala Ile Ala Cys Pro
85 90 95

Cys Gly Gly Ile Gly Thr Arg Gln Phe Pro Ile Val Gln Gln Thr Leu
100 105 110

Gln Leu Leu Pro His Cys Tyr Gln His Lys Gln Ile Asp Ser Ser Arg
115 120 125

Ile Tyr Pro Leu Gln Ile Asn
130 135

<210> 208
<211> 113
<212> PRT
<213> Homo sapien

<400> 208

Met Thr Leu Asn Glu His Ala Ala Phe Lys His Leu Phe Asn Lys Ala
1 5 10 15

His Leu Ala Leu Pro Leu Ile His Leu Thr Leu Ser Gly His Arg Thr
20 25 30

Cys Phe Arg Glu His Arg Val Gly Gly Lys Val Thr Asp Gln Gln Asp
35 40 45

Pro Lys Ala Glu Glu Phe Phe Leu Val Ala Asn Lys Met Lys Ser Leu
50 55 60

Pro Cys Leu Leu Leu Ser Thr Gln Thr Arg Gln Pro Ser Asp Phe Ser
65 70 75 80

Ile Phe Ser Pro Pro Phe Pro Pro Phe Tyr Ser Thr Lys Pro Pro Leu
85 90 95

Ser Ser Trp Pro Val Leu Asn Glu Leu Leu Gly Thr Cys Pro Gly Gly
100 105 110

Arg

<210> 209
<211> 72
<212> PRT

136

<213> Homo sapien

<400> 209

Met Leu Leu Gly Ala Ala Pro Cys Asp Gly Ser Ala Ala Arg Ala Val
1 5 10 15

Val Ile Pro Ala Thr Trp Glu Ala Glu Ala Glu Asn Cys Leu Asn Pro
20 25 30

Gly Gly Arg Gly Cys Ser Glu Ser Arg Ser Tyr His Cys Thr Pro Ala
35 40 45

Arg Ala Thr Glu Gly Asp Ser Ile Ser Lys Lys Arg Lys Lys Gly Lys
50 55 60

Ala Gly Leu Ser Gly Ser His Leu
65 70

<210> 210

<211> 74

<212> PRT

<213> Homo sapien

<400> 210

Arg Pro Ser Ala Val Thr His Ala Cys Asn Pro Ser Thr Leu Gly Gly
1 5 10 15

Gln Gly Gly Trp Ile Thr Arg Ser Gly Asp Gln Asp His Pro Gly Ala
20 25 30

His Gly Glu Thr Pro Ser Leu Leu Lys Ile Gln Lys Ile Ser Pro Val
35 40 45

Trp Trp Trp Ala Pro Val Val Pro Ala Thr Arg Glu Ala Glu Ala Gly
50 55 60

Glu Trp Arg Glu Pro Gly Arg Val Glu Leu
65 70

<210> 211

<211> 71

<212> PRT

<213> Homo sapien

<400> 211

Met Thr Asp Pro Leu Gly Gln Arg Arg Lys Ala Phe Gly Arg Leu Asn

137

1

5

10

15

Ser Asn Arg Ala His Gln Ala Trp Phe Pro Leu Val Val Ala Thr Phe
 20 25 30

Arg Phe Thr Pro Val Ser Pro Ile Val Pro Gln Arg Arg Ile His His
 35 40 45

Leu Glu Ala Thr Pro Thr Arg Arg Phe Lys Val Asp Pro Arg Gly Asp
 50 55 60

Pro Trp His Val Asn Pro Phe
 65 70

<210> 212
 <211> 71
 <212> PRT
 <213> Homo sapien

<400> 212

Met Gln Cys Glu Trp Phe Gln Ile Phe Trp Ser Leu Ser Val Leu Ser
 1 5 10 15

Thr Gln Asn Pro Phe Ser Tyr Pro Cys Leu Ile His Leu Ser Glu Arg
 20 25 30

Thr Tyr Pro Ser Val Leu Lys Tyr Met Tyr Glu His Pro Arg Phe Ser
 35 40 45

Leu Asn Val Trp Ser Ala Phe Ile Thr His Ser Ala Asn Glu Thr Ser
 50 55 60

Pro Ser His Ala Arg Met Leu
 65 70

<210> 213
 <211> 155
 <212> PRT
 <213> Homo sapien

<400> 213

Met Glu Val Gly Ala Val Gly Arg Ser Val Pro Arg Leu Ser Val Phe
 1 5 10 15

Val Leu Leu Ser Arg Arg Ser Val Leu Ser Phe Arg Leu Leu Leu Leu
 20 25 30

138

Phe Val Arg Pro Ser Gly Pro Ser Gly Pro Pro Phe Cys Leu Ser Leu
35 40 45

Ser Leu Leu Ser Val Gly Leu Ser Phe Phe Phe Cys Ser Phe Phe Leu
50 55 60

Ala Phe Pro Gly Pro Cys Thr Val Thr Val Pro Phe Arg Ser Val Ser
65 70 75 80

Val Ser Val Leu Pro Ser Phe Leu Leu Ser Phe Phe Leu Ser Leu Ser
85 90 95

Leu Ser Leu Ser Phe Phe Leu Ser Phe Phe Leu Ser Phe Phe Leu Ser
100 105 110

Phe Phe Leu Ser Phe Phe Leu Gly Ser Cys Ser Val Thr Gln Gly Gly
115 120 125

Glu Arg Trp His His His Ser Leu Leu Gln Ser Gln Leu His Arg Leu
130 135 140

Lys Gln Ser Ser Tyr Leu Ile Val Leu Ser Ser
145 150 155

```
<210> 214
<211> 103
<212> PRT
<213> Homo sapien
```

<400> 214

Phe Phe Leu Ser Phe Ser Gly Leu Ala Leu Ser Pro Lys Val Glu Ser
1 5 10 15

Gly Gly Ile Ile Thr Ala Tyr Cys Ser Leu Asn Phe Thr Gly Ser Ser
20 25 30

Asn Pro Pro Thr Ser Leu Ser Ala Val Ala Glu Thr Ala Gly Met Cys
35 40 45

His His Ala Pro Leu Ile Phe Val Tyr Phe Leu Glu Thr Gly Phe Leu
50 55 60

His Val Ala His Ala Gly Leu Glu Leu Phe Gly Ser Ser Ser Ser Pro
65 70 75 80

139

Ala Ser Ala Ser Gln Ser Ala Arg Ile Thr Gly Val His His Cys Ala
85 90 95

Trp Pro Thr Ala Met Phe Ser
100

```
<210> 215
<211> 125
<212> PRT
<213> Homo sapien
```

<400> 215

Met Asn Ala Ser Thr Asp Asp Thr Leu Thr Gln Glu Trp Gln Pro Ser
1 5 10 15

Thr Asn Ala Ala Leu Ala Asn His Tyr Thr Gly Ser Ser Pro Ser Gly
20 25 30

Gly Arg Leu Ala Leu Pro Leu Ser Asp Glu Leu Ile Leu Gln Arg Asp
35 40 45

Ile Glu Ser Ser Arg Leu Ile Ser Ser Cys Trp Gly Pro Pro Ile Asn
50 55 60

His Ala Gly Ser Pro Arg Tyr Ser Gln Ala Asp Lys Pro Gln Gly Tyr
65 70 75 80

His Thr Arg Phe Gly Glu Val Asn Leu Ser Ala Arg Gly Gly Ser Gly
85 90 95

Ala Cys Leu Asp His Met Val Gln Gly Glu Ala Phe Gln Gly Leu Thr
100 105 110

Gln Leu Ser Arg Gly Arg Ala Cys Thr Ser Ala Ala Thr
115 120 125

```
<210> 216
<211> 76
<212> PRT
<213> Homo sapien
```

<400> 216

Met Gly Val Gly Thr Thr Gln Gly Pro Pro Tyr Lys Ala Gly Phe Phe
1 5 10 15

140

Ser Ile Lys Ser Tyr Thr Lys Val Cys Leu Pro Leu Leu Pro Gly Phe
 20 25 30

Leu His Leu Phe His Pro Leu Leu Thr Ser Gly Ala Gly Lys Thr Lys
 35 40 45

Pro Ser Ser Ser Ser Leu Leu His Ser Leu Leu Ser Ala Lys Thr Val
 50 55 60

Arg Asp Glu Asp Phe Ser Asp Asp Pro Leu Ser Thr
 65 70 75

<210> 217
 <211> 42
 <212> PRT
 <213> Homo sapien

<400> 217

Met Leu Pro Leu Ala Gly Val Gln Trp Tyr Arg Ser Arg His His Cys
 1 5 10 15

Asn Leu Cys Leu Thr Arg Val Gln Ala Asn Ser Leu His Ser Ala Ser
 20 25 30

Gln Val Ala Gly Ile Thr Thr Cys Pro Ala
 35 40

<210> 218
 <211> 82
 <212> PRT
 <213> Homo sapien

<400> 218

Met Ala Trp Leu Gly Leu Arg Gly Leu Thr Phe Leu Pro Ser Tyr Ile
 1 5 10 15

Asn Lys Lys Asn Lys Thr Asn Ser Val Glu Val Leu Gly Trp Gln Lys
 20 25 30

Phe Leu Gly Gly Asp Met Glu Arg Glu Trp Ala Met Phe Leu Arg Ala
 35 40 45

Ala Ser Ser Gly Ile Arg Gly Gly Val Gly Thr Phe His Cys Glu Ser
 50 55 60

141

Tyr Pro Lys Leu Gly Ile Arg Asp Gly Leu Gly Gly Ser Arg Asp Leu
 65 70 75 80

Gly Arg

<210> 219
 <211> 72
 <212> PRT
 <213> Homo sapien

<400> 219

Asp Tyr Ala Glu Ser Pro Ala Ala Leu Ser Asn Gln Thr Ser Ala Val
 1 5 10 15

Val Pro Ile Leu Arg Pro Phe Ile Pro Val Phe Leu Leu Leu Leu Phe
 20 25 30

His Leu Val Phe Gln Phe Ile Gln Asn Arg Ile Gln Ala Ile Thr Asn
 35 40 45

His Ser Ile Ala Gln Met Phe Leu Leu Thr Ser Pro Gln Ser His Pro
 50 55 60

Leu Pro Gln Asp Leu Pro Ser Ala
 65 70

<210> 220
 <211> 65
 <212> PRT
 <213> Homo sapien

<400> 220

Met Ser Gly Gly Gln Arg Glu Arg Leu Asp Thr Gly Glu Gly Gly Asn
 1 5 10 15

Val Thr Thr Ala Ala Arg Cys Tyr Thr Ala Gly Leu Glu Val Glu Glu
 20 25 30

Lys Ala Lys Asn Ala Thr Asn Val Ala Trp Lys Leu Glu Lys Ala Arg
 35 40 45

Lys Leu Phe Ser Leu Arg Thr Ser Gly Gly Ser Val Ala Leu Pro Thr
 50 55 60

His

142

65

<210> 221
 <211> 476
 <212> PRT
 <213> Homo sapien

<400> 221

Lys Met Ser Trp Arg Pro Gln Tyr Arg Ser Ser Lys Phe Arg Asn Val
 1 5 10 15

Tyr Gly Lys Val Ala Asn Arg Glu His Cys Phe Asp Gly Ile Pro Ile
 20 25 30

Thr Lys Asn Val His Asp Asn His Phe Cys Ala Val Asn Thr Arg Phe
 35 40 45

Leu Ala Ile Val Thr Glu Ser Ala Gly Gly Gly Ser Phe Leu Val Ile
 50 55 60

Pro Leu Glu Gln Thr Gly Arg Ile Glu Pro Asn Tyr Pro Lys Val Cys
 65 70 75 80

Gly His Gln Gly Asn Val Leu Asp Ile Lys Trp Asn Pro Phe Ile Asp
 85 90 95

Asn Ile Ile Ala Ser Cys Ser Glu Asp Thr Ser Val Arg Ile Trp Glu
 100 105 110

Ile Pro Glu Gly Gly Leu Lys Arg Asn Met Thr Glu Ala Leu Leu Glu
 115 120 125

Leu His Gly His Ser Arg Arg Val Gly Leu Val Glu Trp His Pro Thr
 130 135 140

Thr Asn Asn Ile Leu Phe Ser Ala Gly Tyr Asp Tyr Lys Val Leu Ile
 145 150 155 160

Trp Asn Leu Asp Val Gly Glu Pro Val Lys Met Ile Asp Cys His Thr
 165 170 175

Asp Val Ile Leu Cys Met Ser Phe Asn Thr Asp Gly Ser Leu Leu Thr
 180 185 190

Thr Thr Cys Lys Asp Lys Lys Leu Arg Val Ile Glu Pro Arg Ser Gly

143

195

200

205

Arg Val Leu Gln Glu Ala Asn Cys Lys Asn His Arg Val Asn Arg Val
 210 215 220

Val Phe Leu Gly Asn Met Lys Arg Leu Leu Thr Thr Gly Val Ser Arg
 225 230 235 240

Trp Asn Thr Arg Gln Ile Ala Leu Trp Asp Gln Glu Asp Leu Ser Met
 245 250 255

Pro Leu Ile Glu Glu Glu Ile Asp Gly Leu Ser Gly Leu Leu Phe Pro
 260 265 270

Phe Tyr Asp Ala Asp Thr His Met Leu Tyr Leu Ala Gly Lys Gly Asp
 275 280 285

Gly Asn Ile Arg Tyr Tyr Glu Ile Ser Thr Glu Lys Pro Tyr Leu Ser
 290 295 300

Tyr Leu Met Glu Phe Arg Ser Pro Ala Pro Gln Lys Gly Leu Gly Val
 305 310 315 320

Met Pro Lys His Gly Leu Asp Val Ser Ala Cys Glu Val Phe Arg Phe
 325 330 335

Tyr Lys Leu Val Thr Leu Lys Gly Leu Ile Glu Pro Ile Ser Met Ile
 340 345 350

Val Pro Arg Arg Ser Asp Ser Tyr Gln Glu Asp Ile Tyr Pro Met Thr
 355 360 365

Pro Gly Thr Glu Pro Ala Leu Thr Pro Asp Glu Trp Leu Gly Gly Ile
 370 375 380

Asn Arg Asp Pro Val Leu Met Ser Leu Lys Glu Gly Tyr Lys Lys Ser
 385 390 395 400

Ser Lys Met Val Phe Lys Ala Pro Ile Lys Glu Lys Lys Ser Val Val
 405 410 415

Val Asn Gly Ile Asp Leu Leu Glu Asn Val Pro Pro Arg Thr Glu Asn
 420 425 430

144

Glu Leu Leu Arg Met Phe Phe Arg Gln Gln Asp Glu Ile Arg Arg Leu
 435 440 445

Lys Glu Glu Leu Ala Gln Lys Asp Ile Arg Ile Arg Gln Leu Gln Leu
 450 455 460

Glu Leu Lys Asn Leu Arg Asn Ser Pro Lys Asn Cys
 465 470 475

<210> 222
 <211> 85
 <212> PRT
 <213> Homo sapien

<400> 222

Met Gly Pro Arg Cys Cys Ser Ser Gly Ala Ser Val Met Asp Glu Arg
 1 5 10 15

Pro Pro Gly Gln Val Val Ile Pro Ala Thr Trp Glu Ala Glu Ala Glu
 20 25 30

Asn Cys Leu Asn Pro Gly Gly Arg Gly Cys Ser Glu Ser Arg Ser Tyr
 35 40 45

His Cys Thr Pro Ala Arg Gln Gln Lys Glu Thr Pro Ser Gln Lys Lys
 50 55 60

Glu Lys Lys Val Arg Pro Asp Ser Val Ala His Thr Cys Asn Leu Ser
 65 70 75 80

Thr Ser Gly Gly Gly
 85

<210> 223
 <211> 75
 <212> PRT
 <213> Homo sapien

<400> 223

Arg Pro Ser Ala Val Ala His Ala Cys Asn Pro Ser Thr Leu Gly Gly
 1 5 10 15

Gln Gly Gly Trp Ile Thr Arg Ser Gly Asp Ala Asp His Pro Gly Ala
 20 25 30

His Gly Glu Thr Pro Ser Leu Leu Lys Ile Gln Lys Ile Ser Pro Val

145

35

40

45

Trp Trp Trp Ala Pro Val Val Pro Ala Thr Arg Glu Ala Glu Gly Gly
 50 55 60

Glu Trp Arg Glu Pro Gly Arg Trp Ser Leu Gln
 65 70 75

<210> 224
 <211> 61
 <212> PRT
 <213> Homo sapien

<400> 224

Met Pro Leu His Ser Ser Leu Gly Asn Ile Val Arg Ser Cys Leu Lys
 1 5 10 15

Asn Asn Asn Asn Lys Ile Gly Arg Ala Arg Trp Leu Thr Pro Val Ile
 20 25 30

Pro Ala Leu Trp Glu Ala Lys Ala Gly Gly Ser Arg Gly Gln Glu Ile
 35 40 45

Lys Thr Ile Leu Ala Asn Thr Val Lys Pro His Leu Tyr
 50 55 60

<210> 225
 <211> 75
 <212> PRT
 <213> Homo sapien

<400> 225

Arg Pro Ser Ala Val Ala His Ala Cys Asn Pro Ser Thr Leu Gly Gly
 1 5 10 15

Gln Gly Gly Trp Ile Thr Arg Ser Gly Asp Gln Asp His Pro Gly Ala
 20 25 30

His Gly Glu Thr Pro Ser Leu Leu Lys Ile Gln Lys Ile Ser Pro Val
 35 40 45

Trp Trp Trp Ala Pro Val Val Pro Ala Thr Arg Glu Ala Glu Ala Gly
 50 55 60

Asp Trp Arg Glu Pro Gly Arg Trp Ser Leu Gln
 65 70 75

146

<210> 226
 <211> 67
 <212> PRT
 <213> Homo sapien

<400> 226

Met Leu Glu Arg Arg Gln Cys Asp Gly Cys Val Val Ala Ala Gly Gly
 1 5 10 15

Thr Ile Lys Thr Glu Gly Glu His Asp Pro Val Thr Glu Phe Ile Gly
 20 25 30

Glu Ala Asp Cys Leu Ala Leu Tyr Tyr Asn Arg Lys Cys Gln Leu Gly
 35 40 45

Ala Val Ala His Ala Cys Asn Pro Ser Thr Leu Gly Gly Gln Gly Gly
 50 55 60

Trp Ile Thr
 65

<210> 227
 <211> 105
 <212> PRT
 <213> Homo sapien

<400> 227

Met His Ala Arg Ala Ala Gln Cys Asp Gly Ser Ala Ala Arg Ala Gly
 1 5 10 15

Thr Cys Trp Arg Arg Glu Thr Thr Arg Thr Ala Ala Ser Leu Gly Pro
 20 25 30

Val Thr Leu Arg Asp Met Asp Glu Ala Gly Asn His His Ser Gln Gln
 35 40 45

Thr Asn Thr Glu Ala Glu Asn Gln Thr Pro His Val Leu Thr His Lys
 50 55 60

Trp Glu Leu Asn Ser Glu Asn Thr Trp Thr Gln Gly Gly Glu His His
 65 70 75 80

Thr Pro Arg Pro Val Arg Glu Trp Gly Thr Arg Gly Arg Glu Ser Met
 85 90 95

147

Gly Gln Ile Pro Asn Ala Cys Thr Ala
100 105

<210> 228
<211> 61
<212> PRT
<213> Homo sapien

<400> 228

Met Asn Thr Thr Leu Arg Ala Ser Tyr Ser Lys Arg Ser Cys Arg Ile
1 5 10 15

Arg Phe Asp Ser Arg His Arg Ser Thr His Gln Ala His Gly Ile Trp
20 25 30

Ala Val His Ser Leu Gly Ser Tyr Val Phe Ala Ser Ser Ser Ala Ala
35 40 45

Ile Leu Ala Ser Pro Gly Ser Ile Asn Ser Cys Ile Lys
50 55 60